# Victimization and Skill Accumulation: The Case of School Bullying*

Miguel Sarzosa†
*Purdue University*

September 19, 2018

**Abstract**

Recent literature has shown that skills are not only essential for the development of successful adults, but also that they are malleable and prone to be affected by many experiences, especially during childhood. This paper examines how bullying depletes skills in school children. I formulate a dynamic model of skill accumulation with endogenous victimization based on the identification of unobserved heterogeneity. I allow victimization to depend on each student's traits and those of her classmates. Using a unique longitudinal dataset on middle school students, I find that victimization depletes current skill levels by 40% of a standard deviation for the average child. This skill depletion causes the individual to become 25% more likely to experience bullying again. Therefore bullying triggers a self-reinforcing mechanism that opens an ever-growing skill gap. Finally, I find evidence that supports the allocation of students in more skill-homogeneous classrooms as a tool to reduce victimization.

**Keywords:** Bullying, non-cognitive skills, skill dynamics.

**JEL codes:** I12, I14, I25, I31

*I would like to thank Sergio Urzua for his invaluable support. I would also like to thank John Ham, John Shea, Sebastián Galiani, Soohyung Lee, Brian Quistorff for their comments on earlier stages of this research project. In addition I would like to thank the participants of the Global Education Forum at Seoul, Republic of Korea, and the seminar participants at the University of Maryland, the Swedish Institute for Social Research of Stockholm University, Purdue University, the Psychology Department BBL of Purdue University, LACEA-LAMES meeting (São Paulo, 2015), Royal Economic Society Annual Conference (Bristol, UK, 2017), Western Economic Association International Meetings (San Diego, 2017), and European Economics Association Meetings (Lisbon, Portugal, 2017). The Web Appendix with supplementary material is available at https://goo.gl/G56a9u

†403 W. State Street, West Lafayette, IN, 47907; phone: (765) 494-4343; fax: (765) 494-9658; email, msarzosa@purdue.edu. The author did not have access to information leading to the identification of individuals. The data analysis was carried out in a secure server.

# 1 Introduction

According to psychologists a bullying victim is a person that is repeatedly and intentionally exposed to injury or discomfort by others in an environment where an imbalance of power exists (Olweus, 1997).[1] Sociologists suggest that bullying thrives in contexts where individuals need to show peer group status (Faris and Felmlee, 2011). Not surprisingly, schools are the perfect setting for bullying. The combination of peer pressure, multidimensional heterogeneity of students, and juvenile lack of self-control, makes schools a fertile ground for bullying. In 2013, 22% of US students ages 12 through 18 reported being victimized in school (National Center for Education Statistics, 2015).

Bullying is very costly. Eleven percent of urban American children miss school every day because of fear of being victimized (Kann et al., 2014). One of every ten students drops out or changes school because of it (Baron, 2016). Homicide perpetrators are twice as likely as homicide victims to have been bullied previously by their peers (Gunnison et al., 2016). Victims are between 2 to 9 times more likely to consider suicide than non-victims (Kim and Leventhal, 2008; Kim et al., 2009). Surprisingly, economic literature has remained mostly silent on the topic. Very little is known about its intermediate costs and long-term consequences. In this paper, I contribute in bridging this gap by exploring the two-way relation between bullying and cognitive and non-cognitive skills accumulation.[2] Namely, how school bullying hampers the development of successful adults by impeding skill accumulation, and the extent

[1] Injury or discomfort can be caused by violent contact, by insults, by communicating private or inaccurate information and by other unpleasant behaviors like exclusion from a group.

[2] *Cognitive skills*, defined as "all forms of knowing and awareness such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving" (APA, 2006), and *non-cognitive skills*, vaguely defined as personality and motivational traits that determine the way individuals think, feel and behave (Borghans et al., 2008), are critical to the development of successful lives. See for example Murnane et al. (1995); Cawley et al. (2001); Heckman and Rubinstein (2001); Duckworth and Seligman (2005); Heckman et al. (2006); Urzua (2008); Saltiel et al. (2017). Although psychologists treat them differently, most related works in economics use the terms *non-cognitive* and *socioemotional* skills interchangeably (Saltiel et al., 2017). Sometimes they are also referred to as *soft* skills (Heckman and Kautz, 2012).

to which cognitive and non-cognitive skills are themselves determinants of in-school victimization.[3] To analyze this two-way relation, I extend the theoretical contributions of Cunha et al. (2010) to include peer-influenced events—like bullying—in the skills accumulation process. I allow future skills to depend on current skills, current investment choices and victimization. I allow the likelihood of the bullying event to depend on own and peer observable and unobservable characteristics. Thus, I treat bullying as an event that may deplete the existing stock of skills changing negatively the skill accumulation path for the people involved.

The model incorporates several desirable features. First, it acknowledges that social interactions like bullying are endogenous. Hence, the "treatment" is not randomly allocated across students. The way own characteristics relate to those of peers is key in building up the social arena that determines victimization. Second, it recognizes that cognitive and non-cognitive skill measures observed by the researcher are only approximations or functions of the true latent skills (Carneiro et al., 2003; Heckman et al., 2006). Third, it does not assume normality of the latent factors' distributions. This guarantees the flexibility required to appropriately recreate the unobserved distributions in the estimation. Finally, the model allows me to simulate counterfactuals for each skill level, which I use to calculate the divergence in skill accumulation paths caused by bullying.

This paper contributes to the economic literature in several ways. First, it extends the literature on dynamic skill accumulation by introducing peer-influenced events as important drivers of the process. Second, it analyzes the consequences of disruptive behaviors in school in terms of skill depletion. Third, it extends my previous work on school bullying (Sarzosa and Urzua, 2015), where I found sizable consequences borne during adulthood, by providing additional insight into the channels through which

[3]This two-way relation is inspired by facts about child victimization that are well established in the psychological literature. Namely, that bullying victims suffer grave and long-lasting consequences in terms of their emotional well being (Olweus, 1997; Smith and Brain, 2000, among many others), and that the likelihood of victimization increases dramatically when the child has some behavioral vulnerability (Hodges et al., 1997).

high school bullying affects adult outcomes. Fourth, it allows the quantification of the long-run cost of bullying. That is, I can go beyond short and medium-term outcomes like school absenteeism or young adult health, and estimate skill endowments loses for life. In addition, this will open an auspicious research agenda on skill accumulation and negative social interactions.

Using detailed longitudinal data on a cohort of South Korean students, I find that kids with low initial stocks of skills and those who have uncommon traits among their peers are significantly more likely to be bullied. I also find that victimization depletes current skill levels and makes individuals more prone to experiencing bullying again in the future, creating a self-reinforcing mechanism that generates a big burden to be carried during adulthood.

This paper is organized as follows. After reviewing the scarce economic literature on the subject in Section 2, I present the basic framework for the analysis of skill dynamics in Section 3. Section 4 defines the empirical strategy I will use in this paper. Section 5 describes the data I use for the analysis, and describes of how the cognitive and non-cognitive skill measures are constructed. Section 6 presents my results. Section 7 focuses on how, in light of my results, some policies regarding students allocation to school can reduce school bullying. Finally, Section 8 concludes.

## 2 Related Literature

Although economic literature on the consequences of disruptive peers has grown in recent years (Carrell and Hoekstra, 2010, 2012; Carrell et al., 2016), economic research on bullying is scarce.[4] Two main reasons explain this sparseness: first, lack

[4]The psychology and sociology literatures have been prolific in examining bullying as a social phenomenon. Among many findings, they have stablished that school and class size are not significant determinants of the likelihood of bullying, nor are personal characteristics like disabilities, obesity, hygiene, posture and dress (Olweus, 1997). However, victims are often smaller than attackers (Olweus, 1997), and victims have more odd mannerisms than non-victimized kids (Lowenstein, 1978). Victims have fewer friends and are more likely to be absent from school (Smith et al., 2004). Bullied children in general have less self-esteem, and have a negative view of their situation (Björkqvist

of adequate data; and second, the non-randomness of the selection into bullying. Regarding the former, there is little longitudinal data that inquires about bullying, so there are very few sources that researches can use to observe individuals before and after the event. Regarding the latter, the non-randomness of bullying causes the consequences of bullying to be confounded by the intrinsic characteristics that made the person a victim or a perpetrator in the first place. The scarce existing economic literature has focused on the quantification of the effects bullying has on short and medium-term outcomes. Brown and Taylor (2008) find that being bullied and being a bully are correlated with lower educational attainment in the UK. Eriksen et al. (2014) find that being bullied decreases $9^{th}$ grade GPA among Danish students. They are able to find causal estimates by instrumenting victimization with the proportion of classroom peers whose parents have a criminal conviction. Sarzosa and Urzua (2015) embed a similar empirical strategy (i.e., instrumenting victimization with the proportion of classroom peers that come from violent families and the proportion who claim to be bullies) in a framework of unobserved heterogeneity in terms of cognitive and non-cognitive skills. They find that bullying increases the probability of smoking, the likelihood of feeling sick, depressed, stressed and unsatisfied with life during adulthood. It also reduces college enrollment and increases the dislike of school. Interestingly, they find that the detrimental effects of bullying are greater for people with low levels of non-cognitive sills.

In this paper, I contribute to the analysis of bullying literature by building on Sarzosa and Urzua (2015) and providing an explanation for how the gaps in the outcomes observed in that paper materialize. That is, while Sarzosa and Urzua (2015) estimate the ATE of middle school bullying on young adult outcomes, in the present study, I elucidate the mechanism behind the creation of these gaps, showing bullying as the triggering event that determines a divergence in skill accumulation

et al., 1982; Olweus, 1997). They are also more likely to feel lonely (Dake et al., 2003). These characterizations of the victims highlight the importance of including explicit relation between bullying and personality throughout the analysis.

paths. In the present paper, I show that the gaps between victims and non-victims open up early in life by embarking victims in skill accumulation paths that are lower than those they would have been on in the absence of victimization. In consequence, the identification strategies in both papers differ greatly. While Sarzosa and Urzua (2015) estimate a static Roy model with unobserved heterogeneity, in the present paper I estimate a dynamic model of skill formation. Although both papers use the allocation of students to classrooms as exogenous variation for identification, in the present study, I use it as an exogenous shifter of how uncommon students' traits are in the pool of traits available in the classroom.

This paper also relates to recent contributions in the area of skill development that have stablished that skills are dynamic and malleable (Cunha et al., 2006, 2010). That is, they depend on their past levels, they can be hindered and they can be fostered. We know that skills beget skills and therefore, initial skill endowments and early accumulation are critical for the lifetime stock of skills (Cunha et al., 2006). This self-reinforcing mechanism increases skill inequality as children age: those who start their childhood with high initial levels of skills accumulate skills three times faster than those who start their development with very low stocks of skills (Agostinelli and Wiswall, 2016a). This gives foundation to the call for early childhood development and preschool interventions (Knudsen et al., 2006; Doyle et al., 2009). Skills beget skills not only though the natural process of getting the stock available at time $t$ to $t+1$, but also though investment. That is, skilled kids receive more skill investment and have higher returns to those investments than less skilled kids (Skinner and Belmont, 1993; Aizer and Cunha, 2012; Espinoza et al., 2014).[5]

The claim that skills are malleable is backed up by a series of papers that show that skill developing interventions were able to modify the stock of skills of the treated population. For instance, the people treated by Perry Preschool Program have higher non-cognitive skills—although similar levels of cognitive skills—than the

[5]Empirical estimates back up the theoretical claim of skills inducing higher levels of investment only at very early stages of life (i.e., before two years of age)(Cunha et al., 2010).

controls (Heckman et al., 2010). The Socio-Emotional Learning programs have been widely reviewed as successful interventions that develop non-cognitive skills such as goal setting, conflict resolution and decision making (Payton et al., 2008). In fact, skill developing interventions can compensate for low initial levels of both cognitive and non-cognitive skills (Cunha et al., 2010). Furthermore, extensive literature finds that family background influences skill accumulation: children whose parents are more engaged in their upbringing are likely to have higher levels of skill.[6] There is evidence, however, that there are windows of opportunity outside of which skill malleability is lost (Knudsen et al., 2006), and that such windows close earlier for cognitive than for non-cognitive skills (Cunha et al., 2006).

Besides the dynamism and malleability features of skills, recent literature has found that they strongly depend on the contexts the child grows in.[7] For instance, skill endowments depend on the level of stress a person was exposed to during childhood (McEwen and Seeman, 2006) and the quality of school inputs such as class size and teacher characteristics (Fredriksson et al., 2013; Jackson, 2013). One of such contexts is the type of social interactions the child encounters in school. This paper includes interactions with peers as critical inputs in the development of own skills.

# 3 Skill Formation With Peer-Influenced Inputs

My framework needs to incorporate five facts that emanate from the skill formation literature: i. skills beget skills, ii. skill development can be affected by investment choices, iii. past skills levels can affect next period skills indirectly by inducing skill investment, iv. social interactions like bullying can hamper skill development, and v. the nature of those social interactions depends also on the stock of cognitive and non-cognitive skills of each person and those of his or her peers. Therefore,

[6] See, for instance, Hart and Risley (1995); Cunha et al. (2006); Heckman and Masterov (2007); Cabrera et al. (2007); Kiernan and Huerta (2008); Tough (2012); Attanasio et al. (2017).

[7] See OECD (2014) for a full framework about such contexts.

I propose to augment the dynamic structure in Cunha et al. (2010) to explicitly incorporate these five facts. Let the stock of skills $S = \{A, B\}$ a person $i$ that belongs to classroom $c \in C$ has at time $t+1$ (i.e., $\theta_{S,i\in c,t+1}$) be a result of a CES skill production function whose inputs are the stock of skills she had at time $t$ ($\theta_{A,i\in c,t}$ and $\theta_{B,i\in c,t}$), and the skill investment choices done between the two periods ($I_{S,i\in c,t}$).[8] In addition, let the parameters that define the production function be dependent on the victimization status ($M_{i\in c,t}$). Furthermore, I allow for parental investment choices and the bullying occurrence to be affected by the previous levels of skills. This structure relates parental investment with victimization at $t+1$ indirectly through its effect on the stock of skills at $t$. This relies on the results of psychological research that indicates that responsive and supporting parenting practices are related with lower levels of bullying (Flouri and Buchanan, 2002). In particular, certain parental behaviors that hamper the development of locus of control on kids have been linked with in-school victimization (Ladd and Ladd, 1998).

Hence, the model of skill formation that allows for endogenous peer-influenced events can be described through the following equations:

$$\theta_{S,i\in c,t+1} = \left[\gamma^{M_i}_{A,S,t}\theta^{\rho^{M_i}_S}_{A,i\in c,t} + \gamma^{M_i}_{B,S,t}\theta^{\rho^{M_i}_S}_{B,i\in c,t} + \gamma^{M_i}_{I,S,t}I^{\rho^{M_i}_S}_{S,i\in c,t+1}\right]^{1/\rho^{M_i}_S} + \eta^{M_i}_{S,i\in c,t} \tag{1}$$

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

$$M_{i\in c,t+1} = \mathbf{1}\left[\mathbf{X}_{it}\beta^M_{t+1} + \alpha^{M_A}_{t+1}\theta_{A,i\in c,t} + \alpha^{M_B}_{t+1}\theta_{B,i\in c,t} + \Lambda^{M_c}_{t+1}\nabla_{\psi,i\in c}(d) + \Gamma Z_c > e^M_{i\in c,t+1}\right] \tag{2}$$

for $S = \{A, B\}$, where $\gamma^{M_i}_{I,S,t} = 1 - \gamma^{M_i}_{A,S,t} - \gamma^{M_i}_{B,S,t}$, $\mathbf{1}[\cdot]$ is an indicator function that takes the value of 1 if true and $-i \in c$ indicates all individuals that belong to classroom

[8] The choice of the CES as the production function of skills responds to two main reasons. First, it follows the existing literature, that way I can rely on some identifying assumption already outlined in Cunha et al. (2010). Second, it provides the curvature needed to explore complementarities between the inputs involved in the skill production function. In particular, the *static complementarity* $\left(\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial\theta_{S,t}\right)$ and the *dynamic complementarity* $\left(\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial I_{S,t}\right)$, concepts introduced by Cunha and Heckman (2008) to describe how the current stock of skills affect the productivity of skill investment, and how much of that productivity of investment is leveraged by past investment choices. I will use the same concepts to analyze the *skill depleting* power of the bullying event.

$c$ except $i$. I assume that $\eta^{M_i}_{S,i\in c,t}$ , $\varepsilon_{S,i\in c,t+1}$, and $e^{M}_{i\in c,t+1}$ are *iid* shocks orthogonal to contemporaneous skills, to each other, across time and across victimization condition. Furthermore, I assume that $e^{M}_{i\in c,t+1} \sim \mathcal{N}\left(0, \sigma_{e^{M}_{t+1}}\right)$.

Through the victimization equation (2), I incorporate two stylized facts of bullying established by the psychological literature: i. that there are personal characteristics of the student that influences the chances of being bullied (i.e., behavioral issues), and ii. that there are characteristics of the peer group that set her apart from her classmates (e.g., lacks friends, is rejected by the peer-group) (Hodges et al., 1997). The victimization equation responds to the fact that bullying needs a social arena in which the imbalances of power take place allowing classmates to play different roles: victim, perpetrator and bystanders.[9] Therefore, the question that arises is: what separates bystanders from victims? Due to its social setting, one may be inclined to look for answers to this question in the social interactions literature as in Schelling (1971), Pollak (1976) and Manski (1993), where agents interact through their decisions. The problem with bullying is that no one *decides* to be a victim. Hence, while the social interactions literature explains "why do members of the same group tend to behave similarly" (Manski, 2000), I am instead interested in answering *why is this kid chosen among the rest*. "Selection into bullying" is non-random and, like in a social interactions situation, it depends on characteristics of the victim and its classmates, but in a very different way. The idea is that individual $i$ with skills set $(\theta_{A,i,t}, \theta_{B,i,t})$ and observable traits $\mathbf{X}_{it}$ might be bullied in classroom $c$ but not in classroom $c'$. This difference depends on the distributions of skills and traits that the other students bring to each classroom. Therefore, I model they way classmates' traits affect student $i$'s probability of being bullied in a given classroom by introducing a measure of how rare within that classroom the potential victim's traits are. This uncommonness feature is important because it operationalizes the imbalance of power bullying

[9]Psychology literature has identified six types of classmates: ringleader bullies, follower bullies, reinforcers, defenders, bystanders and victims (Salmivalli et al., 1996). Due to data and computational restrictions, I compress the types of classmates to three: bullies, bystanders and victims.

requires based on the fact that kids with uncommon characteristics are more easily regarded as weird and unlikeable, which fosters peer rejection (Hodges et al., 1997).[10]

I measure uncommonness by counting the number of classmates that lie inside an epsilon-ball in the skills and income space that is defined around those qualities for every kid. The intuition is that if your characteristics set you apart, meaning there are no kids similar to you (i.e., low count in your epsilon-ball), you may have higher chances of being bullied. So, $\nabla_{\psi,i\in c}(d)$ is the number of classmates of $i$ in classroom $c$ that lie in an epsilon-ball with radius $d$ in the space of characteristics $\psi$. $Z_c$ contains school or school district characteristics like teachers quality, overall faculty tolerance to bullying, or prevalence of domestic violence in the community that influence the overall likelihood of bullying victimization (Dake et al., 2003). Identification of (2) within model (1) relies on the assumption that the allocation of individual $i$ to classroom $c$ was exogenous, and therefore the assignment of $i$'s classmates is orthogonal to her own traits. As I describe in greater detail in Section 5, I estimate the model using data from South Korean middle schools. The South Korean context is perfect for the identification of (2) thanks to a law that requires school districts to randomly assign students to middle schools and prohibits the grouping of students by ability and achievement levels (Kang, 2007).

The way I introduce classmates' characteristics into the victimization likelihood through $\nabla_{\psi,i\in c}(d)$ is also econometrically advantageous as it goes around the well known problem of peer-effect identification. According to Angrist (2014), randomness in peer allocation is not sufficient to identify peer-effects. He claims that, in order to prevent the unwanted existence of mechanical statistical forces that create spurious correlations, the econometrician needs some observations within the group not to be affected or "treated" by the same peer-effect. In my approach, the uncommonness measure allows for a different "treatment" for every observation to the point that, although everyone is affected by what happens inside their particular epsilon-ball,

[10] Dake et al. (2003) show that students that scored higher on a scale of social acceptance were less likely to be bullied by their peers.

the relative position of those classmates that do not fall within it is irrelevant.

# 4 Empirical Strategy

## 4.1 Measurement System and Unobserved Heterogeneity

The key feature of the empirical strategy is the way it deals with the fact that underlying cognitive and non-cognitive skills and investment preferences are latent rather than observable.[11] They are not well defined entities with measurement scales and instruments, like height and weight are. Instead, these latent constructs need to be inferred from scores, called manifest variables, that can be directly observed and measured (Bartholomew et al., 2011). I start from the assumption of a linear relation between the manifest and the latent variables. It can be thought of as a production function of manifest scores, whose inputs include both the individual observable characteristics and the latent endowments. The empirical strategy incorporates the fact that the observed manifest values respond not only to the latent variables of interest $(\mathbf{\Theta}, I)$, but also to observable traits $(\mathbf{X})$ and random shocks $(e^T, \nu^S)$.

Suppose we follow individuals for two time periods: $t$ and $t+1$. Then, the measurement system is the following:

$$\mathbf{T}_t = \mathbf{X}_{t,T}\beta_t^T + \Lambda_t^T \mathbf{\Theta}_t' + \mathbf{e_t^T} \tag{3}$$

$$\mathbf{T}_{t+1} = \mathbf{X}_{t+1,T}\beta_{t+1}^T + \Lambda_{t+1}^T \mathbf{\Theta}_{t+1}' + \mathbf{e_{t+1}^T} \tag{4}$$

$$\mathcal{I}_{A,t+1} = \mathbf{X}_{t+1,\mathcal{I}}\beta_{t+1}^{\mathcal{I}_A} + \alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{A}} I_{A,t+1} + \nu_{\mathbf{t+1}}^{\mathbf{A}} \tag{5}$$

$$\mathcal{I}_{B,t+1} = \mathbf{X}_{t+1,\mathcal{I}}\beta_{t+1}^{\mathcal{I}_B} + \alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{B}} I_{B,t+1} + \nu_{\mathbf{t+1}}^{\mathbf{B}} \tag{6}$$

$$M_{t+1} = \mathbf{1}\left[\mathbf{X}_{t+1,M}\beta_{t+1}^M + \Lambda_{t+1}^M \Theta_{i\in c,t}' + \Lambda_{t+1}^{M_c}\nabla_{\psi,i\in c}(d) + \Gamma Z_{t+1,c} > e_{t+1}^M\right] \tag{7}$$

[11] In this paper I use the terms *latent variables* and *unobserved heterogeneity* interchangeably. While the term *latent variables* is widely used in statistics, the literature in labor economics prefers the term *unobserved heterogeneity* to differentiate it from the latent variable models that give the basis of probits, logits, censored and truncated estimations.

where $\mathbf{T}_\tau$ is a $L \times 1$ vector that contains the test scores at time $\tau = \{t, t+1\}$, $\mathcal{I}_{S,t+1}$ is a $L_{\mathcal{I}_S} \times 1$ vector that contains each of the investment measures made in skill $S = \{A, B\}$ at time $t+1$. The latent variables of interest are skills $\mathbf{\Theta}_\tau = \begin{bmatrix} \theta_\tau^A & \theta_\tau^B \end{bmatrix}$ and investments $I_{A,t+1}$ and $I_{B,t+1}$. $\mathbf{X}_{\tau,T}$ are matrices with all observable controls affecting test scores at time $\tau = \{t, t+1\}$, and $\mathbf{X}_{t+1,\mathcal{I}}$ is a matrix containing all observable controls affecting manifest investment measures at time $t+1$. $\Lambda_\tau^T$ are loadings matrices of the unobserved skills, while $\alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{A}}$ and $\alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{B}}$ are the same for the unobserved investment factors. I assume that after controlling for observable and unobservable traits, error terms $\mathbf{e}_\tau^\mathbf{T}$ and $\nu_{t+1}^S$ are orthogonal to each other, across time and across equations. That is, I assume that $(\mathbf{\Theta}_\tau, \mathbf{X}_{\tau,T}) \perp \mathbf{e}_\tau^\mathbf{T}$ and that all the elements of the $L \times 1$ vector $\mathbf{e}_\mathbf{t}^\mathbf{T}$ follow a multivariate normal distribution $\mathcal{N}(0, \Sigma_L)$, where $\Sigma_L$ is a $L \times L$ matrix with zeroes in its off-diagonal elements. Likewise, I assume $(I_{A,t+1}, \mathbf{X}_{t+1,\mathcal{I}}) \perp \nu_{\mathbf{t+1}}^\mathbf{A}$ and $(I_{B,t+1}, \mathbf{X}_{t+1,\mathcal{I}}) \perp \nu_{\mathbf{t+1}}^\mathbf{B}$, and that $\nu_{t+1}^S \sim \mathcal{N}\left(0, \Sigma_{L_{\mathcal{I}_S}}\right)$, where $\Sigma_{L_{\mathcal{I}_S}}$ is a square matrix with zeroes in its off-diagonal elements. Furthermore, $\mathbf{e}_\mathbf{t}^\mathbf{T} \perp \mathbf{e}_\mathbf{t+1}^\mathbf{T}$, $\nu_{\mathbf{t+1}}^\mathbf{A} \perp \nu_{\mathbf{t+1}}^\mathbf{B}$ and $\mathbf{e}_\tau^\mathbf{T} \perp \nu_{\mathbf{t+1}}^\mathbf{S}$.

Let me first focus on the identification of $\hat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$ and $\hat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$ from (3) and (4). Given the assumptions made, identification of (5) and (6) and, in particular, the latent investment distributions $\hat{F}_{I_{A,t+1}}(\cdot)$ and $\hat{F}_{I_{B,t+1}}(\cdot)$ is a special case of the identification problems in (3) and (4). In what follows, I describe identification of $\hat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$ and the parameters in (3). Identification of $\hat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$ and the parameters in (4) follow the same intuition.

First, note that the diagonal elements of the matrix $COV(\mathbf{T}_t | \mathbf{X}_{t,T})$ are of the form:

$$COV(T_{t,i}, T_{t,i} | \mathbf{X}_{t,T}) = \left(\alpha_t^{T_i,A}\right)^2 \sigma^2_{\theta_t^A} + \alpha_t^{T_i,A}\alpha_t^{T_i,B}\sigma_{\theta_t^A\theta_t^B} + \left(\alpha_t^{T_i,B}\right)^2 \sigma^2_{\theta_t^B} + \sigma^2_{e_t^{T_i}} \quad (8)$$

and its off-diagonal elements are of the form:

$$COV\left(T_{t,i}, T_{t,j}\left|\mathbf{X}_{t,T}\right.\right) = \alpha_t^{T_i,A}\alpha_t^{T_j,A}\sigma^2_{\theta_t^A} + \left(\alpha_t^{T_i,A}\alpha_t^{T_j,B} + \alpha_t^{T_i,B}\alpha_t^{T_j,A}\right)\sigma_{\theta_t^A\theta_t^B} + \alpha_t^{T_i,B}\alpha_t^{T_j,B}\sigma^2_{\theta_t^B} \tag{9}$$

where $\alpha_t^{T,\cdot}$ are the elements of $\Lambda_t^T$. 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 poses the need of 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 $COV\left(\mathbf{T}_t\left|\mathbf{X}_{t,T}\right.\right)$ 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.[12] 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 $\alpha_t^{T_v,B} = 0$ for $v = \{1, 2, \ldots, L_A\}$ and $L_A > 2$).[13] Therefore, I can organize measurement system (3) such that the subset of measures affected only by $\theta^A$ 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

[12]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.

[13]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 $\Lambda_t^T$ with $\alpha_t^{T_5,A} = 0$ and $\alpha_t^{T_6,A} = 0$. 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).

system in two blocks

$$\begin{bmatrix} \mathbf{T}_t^A \\ \mathbf{T}_t^{A,B} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{t,T}\beta^T + \alpha_t^{(A)A}\theta_t^A + \mathbf{e}_t^{T^A} \\ \mathbf{X}_{t,T}\beta_t^T + \alpha_t^{(A,B)A}\theta_t^A + \alpha_t^{(A,B)B}\theta_t^B + \mathbf{e}_t^{T^{A,B}} \end{bmatrix} \tag{10}$$

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

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

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.[14] It is easy to see that once the $L_A$ loadings are identified, $\sigma_{\theta_t^A}^2$ is also identified.

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

$$\begin{aligned} COV\left(T_{t,m}^{A,B}, T_{t,n}^{A,B} \,|\mathbf{X}_T\right) &= \alpha_t^{(A,B)_m A}\alpha_t^{(A,B)_n A}\sigma_{\theta_t^A}^2 + \alpha_t^{(A,B)_m B}\alpha_t^{(A,B)_n B}\sigma_{\theta_t^B}^2 \\ &\quad + \left(\alpha_t^{(A,B)_m A}\alpha_t^{(A,B)_n B} + \alpha_t^{(A,B)_m B}\alpha_t^{(A,B)_n A}\right)\sigma_{\theta_t^A,\theta_t^B} \end{aligned} \tag{11}$$

$$COV\left(T_{t,m}^{A,B}, T_{t,k}^A \,|\mathbf{X}_T\right) = \alpha_t^{(A,B)_m A}\alpha_t^{(A)_k A}\sigma_{\theta_t^A}^2 + \alpha_t^{(A,B)_m B}\alpha_t^{(A)_k A}\sigma_{\theta_t^A,\theta_t^B}$$

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.[15] 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

[14] 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 $\hat{F}_{I_{A,t+1}}(\cdot)$ and $\hat{F}_{I_{B,t+1}}(\cdot)$ follows the logic used to identify the parameters in the first block of measures in (10).

[15] Unknowns: two loadings per measure minus one that is normalized, $\sigma_{\theta^B}^2$ 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.

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 dynamics governing the production of factor endowments at a given point in time. In particular, in a dynamic and intertwined process in which $\theta_{S,t+1} = g_S\left(\theta_t^A, \theta_t^B\right)$ for $S = \{A, B\}$, $\theta_{t+1}^A \not\perp \theta_{t+1}^B$ holds because of common past influences. That is, $\theta_{t+1}^A$ and $\theta_{t+1}^B$ are correlated because both share common inputs $\theta_t^A$ and $\theta_t^B$, 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 $\Lambda_t^T$: assume that there is one measure among $\mathbf{T}_t^{A,B}$ that is exclusively affected by the second factor (i.e., $\alpha_t^{(A,B)_o A} = 0$ for $o \in [L_A + 1, L_B]$). For presentation simplicity let $T_{t,o}^{A,B}$ also contain the normalized loading for the second factor (i.e., $\alpha_t^{(A,B)_o B} = 1$ ). Then,

$$COV\left(T_{t,l}^A, T_{t,o}^{A,B} \,|\mathbf{X}_T\right) = \sigma_{\theta_t^A,\theta_t^B}$$

$$COV\left(T_{t,m}^{A,B}, T_{t,o}^{A,B} \,|\mathbf{X}_T\right) = \alpha_t^{(A,B)_m B}\sigma_{\theta_t^B}^2 + \alpha_t^{(A,B)_m A}\sigma_{\theta_t^A,\theta_t^B} \tag{12}$$

$$COV\left(T_{t,m}^{A,B}, T_{t,l}^A \,|\mathbf{X}_T\right) = \alpha_t^{(A,B)_m A}\sigma_{\theta_t^A}^2 + \alpha_t^{(A,B)_m B}\sigma_{\theta_t^A,\theta_t^B} \tag{13}$$

for $m = L_A + 1, \ldots, L_B - 1$ and $m \neq o$. Using (13), I can write $\alpha_t^{(A,B)_m A}$ as a function of $\alpha_t^{(A,B)_m B}$ and together with (12), I can write $\sigma_{\theta_t^B}^2$ as a function of $\alpha_t^{(A,B)_m B}$, 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 $\alpha_t^{(A,B)_n B}$ as a function of $\alpha_t^{(A,B)_m B}$ that can be then replaced in (11) 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 $\theta^A$, $\theta^B$ 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).[16]

[16]The Kotlarski Theorem states that if there are three independent random variables $e_{T_1}$, $e_{T_2}$

I do not impose normality to the distribution of the factors $f_{\theta^A,\theta^B}(\cdot,\cdot)$. Instead, I use the mixture of normals in order to achieve the flexibility required to mimic the true underlying distributions of the latent endowments (Attanasio et al., 2017). The mixture of normals not only grants flexibility in the type of distribution it is able to replicate, but also allows numerical integration using the Gauss-Hermite quadrature (Judd, 1998). Numerical integration based on the estimated distribution of the factors is required throughout the whole estimation procedure due to the unobservable nature of the factors. Then, using a Maximum Likelihood estimator, I obtain $\hat{\beta}_t^T, \hat{\Lambda}_t^T$, $\hat{\Sigma}_{L_t}$ and $\hat{F}_{\theta_t^A,\theta_t^B}(\cdot,\cdot)$.

Finally, consider equation (7) which is the empirical version of equation (2). It collects the facts that victimization not only depends on the potential victim's characteristics, but also on the social arena the student faces (i.e., the traits that her classmates bring to the group). As explained in Section 3, I introduce this feature in the model by creating a measure of how uncommon the traits of a given student are among her classmates. To empirically identify such social process, I require the students—and therefore their traits—allocation to classrooms be as good as random. That way, the social arena each student faces is random, and therefore the differences in the probability of being victimized given her traits depends on the differences in the traits' distributions across classrooms. In the same way and as additional exclusion restrictions for the identification of (7), I follow Sarzosa and Urzua (2015) and introduce two additional traits of the social arena of the classroom: the proportion of peers that report being bullies in the class and the proportion of peers in the classroom that come from a violent family.[17]

and $\theta$ 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 $\theta$, up to one normalization. Note that, given that we have already identified all the loadings, we can write (3) in terms of $T_\tau = \theta + e_{T_\tau}$ by dividing both sides by the loadings. See more details in Carneiro et al. (2003).

[17]See the details of these two variables in Sarzosa and Urzua (2015). The *family violence* measure comes from the following questions: 1. I always get along well with brothers or sisters, 2. I frequently see parents verbally abuse each other, 3. I frequently see one of my parents beat the other one, 4. I am often verbally abused by parents, and 5. I am often severely beaten by parents. Answers were aggregated and considered as peers that come from violent families those who have scores above

## 4.2 Dynamic Estimation

### 4.2.1 Identification and Estimation Steps

As shown in the previous section, we can use equations (3) to identify $\hat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$, and equations (5) and (6) to identify $\hat{F}_{I_{A,t+1}}(\cdot)$ and $\hat{F}_{I_{B,t+1}}(\cdot)$. Also, we can use (4) to identify $\hat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$ and consistently estimate $\hat{\Lambda}^T_{t+1}$ and $\hat{\beta}^T_{t+1}$. In consequence, I am able to construct the vector

$$\hat{\xi}_{t+1} = \mathbf{T}_{t+1} - \mathbf{X}_{t+1,T}\hat{\beta}^T_{t+1} = \hat{\Lambda}^T_{t+1}\mathbf{\Theta}'_{t+1} + \mathbf{e^T_{t+1}} \tag{14}$$

Taking advantage of the orthogonality and mutually independence between $\mathbf{e^T_t}$, $\mathbf{e^T_{t+1}}$ and $\eta_{\mathbf{t}}$, and of the non-linearity of the skills production functions, I substitute them from (1) into the measurement system for $\hat{\xi}_{t+1}$. For the sake of brevity, let me call $g^M_{S,t+1}(\theta_t, I_t)$ the production function of skill $S$ at time $t+1$ for those whose victimization condition is $M = \{0,1\}$. Then, I can write (14) as

$$\hat{\xi}_{t+1} = \hat{\alpha}^{\mathbf{T_A}}_{\mathbf{t+1}} g^M_{A,t+1}(\theta_t, I_t) + \hat{\alpha}^{\mathbf{T_B}}_{\mathbf{t+1}} g^M_{B,t+1}(\theta_t, I_t) + \vartheta^M_{t+1}$$

where $\vartheta^M_{t+1} = \hat{\alpha}^{\mathbf{T_A}}_{\mathbf{t+1}}\eta^M_{A,t} + \hat{\alpha}^{\mathbf{T_B}}_{\mathbf{t+1}}\eta^M_{B,t} + \mathbf{e^T_{t+1}}$ is a compounded error term with $E\left[\vartheta^M_{t+1}\right] = 0$ and $var\left[\vartheta^M_{t+1}\right] = \Omega^M_{t+1}$ whose diagonal elements are of the form $\left(\hat{\alpha}^{T^l_A}_{t+1}\right)^2 \sigma^2_{\eta^M_{A,t}} + \left(\hat{\alpha}^{T^l_B}_{t+1}\right)^2 \sigma^2_{\eta^M_{B,t}} + \sigma^2_{e^{T^l}_{t+1}}$ and its off-diagonal elements are of the form $\hat{\alpha}^{T^l_A}_{t+1}\hat{\alpha}^{T^j_A}_{t+1}\sigma^2_{\eta^M_{A,t}} + \hat{\alpha}^{T^l_B}_{t+1}\hat{\alpha}^{T^j_B}_{t+1}\sigma^2_{\eta^M_{B,t}}$. It is straightforward to see that $\Omega^M_{t+1}$ is identified from the fact that $\hat{\Lambda}^T_{t+1}$ and $\hat{\sigma}^2_{\mathbf{e^T_{t+1}}}$ are known from the first stage. Hence, I am effectively reducing the dimensionality of the computational task of estimating the dynamic model. It is now a four-dimensional unobserved heterogeneity problem: two dimensions of skills at $t$ and the investment latent factor for each skill.

the mean. This variable is somewhat similar to the classroom proportion of incarcerated parents variable used as instrument by Eriksen et al. (2014) in that it relates household emotional trauma with violent behavior in school as in Carrell and Hoekstra (2010).

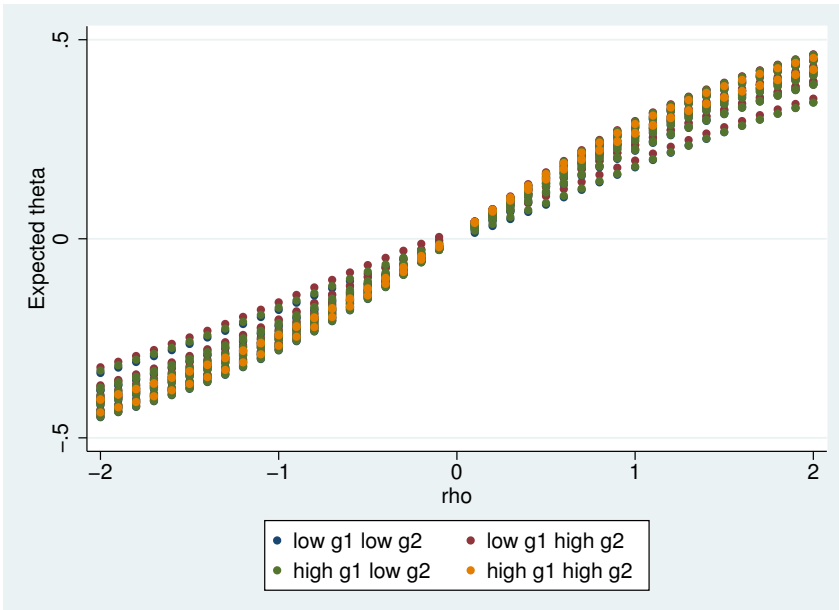
#### 4.2.2 Overall mean shifts and the identification of the CES function

As in Cunha et al. (2010), my estimated factors' distributions are centered at zero. In particular, $E[\theta_{S,t}] = E[\theta_{S,t+1}] = 0$ for $S = \{A, B\}$. These normalizations are at odds with the fact that $E\left[\hat{\theta}_{\cdot,t+1}\right]$ shifts with changes in $\rho$, as shown in Figure 1. It simulates 1,440 different combinations of $\gamma_1$, $\gamma_2$ and $\rho$ to generate $\hat{\theta}_{t+1} = \left[\gamma_1 x^\rho + \gamma_2 y^\rho + (1 - \gamma_1 - \gamma_2) z^\rho\right]^{1/\rho}$, where $x$, $y$ and $z$ come from 5,000 random draws from independent normal distributions. Figure 1 evidences that estimating a model that fits $E\left[\hat{\theta}_{\cdot,t+1}\right] = 0$ greatly constrains the set of possible values that $\hat{\rho}$ can take, and its combinations with the other parameters in the CES function. In other words, the normalizations limit the families of functions that can be estimated. This is consistent with the argument in Agostinelli and Wiswall (2016b) who point out that the normalizations bias the estimations towards finding a functional form consistent with a Cobb-Douglas. One way to fix this is to depart from the estimation procedure put forth by Cunha et al. (2010) as proposed by Agostinelli and Wiswall (2016a). Another way is to use the fact that the relation between $E\left[\hat{\theta}_{\cdot,t+1}\right]$ and the CES parameters is predictable as evidenced by Figure 1. In fact, Table B.1 in the Appendix shows that a flexible cubic polynomial in the CES parameters (i.e., $\mathbf{P}_3(\gamma_1, \gamma_2, \rho)$) captures 99.98% of the variation of $E\left[\hat{\theta}_{\cdot,t+1}\right]$. Hence, in order to avoid $E\left[\hat{\theta}_{\cdot,t+1}\right] = 0$ constraining the possible values of the CES parameters, I use $\mathbf{P}_3(\gamma_A, \gamma_B, \rho_\cdot)$ as a shifter of the mean of $\hat{\theta}_{\cdot,t+1}$ during estimation. That way, it counters the mean-shifting that mechanically occurs when $\hat{\rho} \neq 0$. In practice, I am allowing $E\left[\vartheta^M_{t+1}\right] = -\hat{\alpha}^{T_A}_{t+1}\hat{\mathbf{P}}_3\left(\gamma^M_{A,A}, \gamma^M_{A,B}, \rho^M_A\right) - \hat{\alpha}^{\mathbf{T_B}}_{\mathbf{t+1}}\hat{\mathbf{P}}_3\left(\gamma^M_{A,B}, \gamma^M_{B,B}, \rho^M_B\right)$.

The second implication of the normalizations is that the parameters estimated from (1) will not respond to the overall mean changes in skills. However, given that I am comparing the skill trajectories of victims with those of non-victims, being unable to directly measure overall mean shifts is an innocuous feature of the empirical strategy.[18]

[18] Urzua (2008) shows that—under mild linearity assumptions in measurement systems (3) and

Figure 1: Relation Between the Mean of $\hat{\theta}_{t+1}$ and $\rho$




Note: The $\hat{\theta}_{t+1}$ plotted are the results of 1,440 different combinations of $\gamma_1$, $\gamma_2$ and $\rho$ parameters in the CES production function $\hat{\theta}_{t+1} = \left[\gamma_1 x^{\rho} + \gamma_2 y^{\rho} + (1 - \gamma_1 - \gamma_2)\, z^{\rho}\right]^{1/\rho}$, where $x$, $y$ and $z$ come from 5,000 random draws from independent normal distributions.

### 4.2.3 The problem of joint causality

The empirical model presented so far relies on the assumption that test scores at $t$ are measured before any victimization has occurred. However, given that the data I use in this paper was collected during the school year, there is a chance test scores were measured after some victimization already happened. This may cause a problem of joint causality analogous to the one addressed by Hansen et al. (2004) when exploring the relation between skills, test scores and schooling at the time of measurement. They face the simultaneity issue because schooling is believed to develop skills and boost test scores, but also high skilled people find it easier to achieve higher schooling attainment. Hansen et al. (2004) show that by recognizing that both schooling and

(4)—the mean of the skills is given by the constant terms in $\beta_{\tau}^{T^{A_l}}$ and $\beta_{\tau}^{T^{A_o}}$, call them $\beta_{\tau}^{T^{A_l}}[1]$ and $\beta_{\tau}^{T^{A_o}}[1]$ for $\tau = \{t, t+1\}$. Therefore, I can retrieve overall mean changes of skills from the difference between these constants. For instance, an overall mean change of skill $A$ between $t$ and $t+1$ is given by $\beta_{t+1}^{T^{A_l}}[1] - \beta_{t}^{T^{A_l}}[1]$. Agostinelli and Wiswall (2016a) make use of a similar result to show that a model like (1) can be identifies without normalizing $E\left[\theta_{S,t+1}\right] = 0$.

test scores are generated by common unobserved skills, they can overcome the joint causality problem.

Their approach is well suited for the setting I explore in this paper as it is easy to imagine that—given classmates' traits—both victimization and and test scores observed in the same survey wave are generated by initial unobserved skills. Therefore, using Hansen et al. (2004) framework, I am able to disentangle skills, test scores and victimization. In order to do that, I will extend the structure of the measurement system in (3) to incorporate the one proposed by Hansen et al. (2004).[19]

Let $\mathbf{T}(M_t)$ denote the observed test score at time $\tau$ that depends on the person's victimization condition at the time of the measurement

$$\begin{aligned} \mathbf{T}(M_t) &= \mathbf{X}_{t,T}\beta^T(M_t) + \Lambda^T(M_t)\,\mathbf{\Theta}'_t + \mathbf{e}^{\mathbf{T}}(M_t) \\ M_t &= \mathbf{1}\left(\mathbf{X}_{t,M}\beta^M_t + \Lambda^M_{\tau t}\Theta'_{i\in c,t} + \Lambda^{M_c}_t \nabla_{\psi_t, i\in c}(d) + \Gamma Z_{t,c} > e^M_t\right) \end{aligned} \tag{15}$$

Not that this implies that the matrices $\beta^T_t$ and $\Lambda^T_t$ are expanded to incorporate victimization-dependent coefficients. Also note that this structure is relevant only for the identification of the initial level of skills. For $\tau > t$, the structure of the measurement system remains as in (4).

# 5 Data and Institutional Context

I empirically estimate the described model using the Junior High School Panel (JHSP) of the Korean Youth Panel Survey (KYP). This choice is motivated by two main reasons: South Korea's framework for allocating students to classrooms, and critical

[19]Its identification requires two additional assumptions. First, the assumption of separability between the observed and unobserved part in every equation of the measurement system. Second, the assumption of orthogonality across the error terms in the complete measurement system. The first assumption is trivial given the set up of the empirical model. The second one is a very mild condition as every equation is being controlled not only for observable characteristics but also for the unobserved heterogeneity, which is theorized to be the only source of non-zero covariance between the unobservable parts of all the equations that comprise the complete measurement system.

data features available in the KYP-JHSP.

As explained in Section 3, identification relies on the exogenous assignment of classmates. South Korea's educational setting allows for that thanks to a 1969 "leveling policy" that was introduced to regulate student placement. The law "requires that elementary school graduates be randomly (by lottery) assigned to middle schools—either public or private—in the relevant residence-based school district" (Kang, 2007). The leveling policy also makes the grouping of students by ability and achievement levels "extremely rare". Therefore, the "non-grouping (or ability mixing) in school exposes students to a classroom peer group that is nearly exogenously and randomly determined" (Kang, 2007). Furthermore, the reader should note that unlike in the US, middle-school students in South Korea have a fixed classroom—and hence, classmates—for all subjects.

On top of this distinctive institutional feature, I take advantage of the fact that the KYP-JHSP's has a sampling scheme that is critical for the identification of the peer interactions that fuel the model. The data consist of a nationally representative sample of a cohort of middle schoolers interviewed for the first time in 2003, when they were 14 years old. The importance of the sampling scheme relies on the fact that its sampling unit is the entire classroom. Hence, the KYP-JHSP permits a thorough inspection of the complete distribution of traits available in the classroom, a critical feature for the identification of equation (2). The panel consists of 3,449 youths (see descriptive statistics in Table 1). Subjects were consistently interviewed in six waves; one each year.[20] Each wave, information was collected in two separate questionnaires: one for the teenager, and another one for the parents or guardians.

Another key feature or the KYP-JHSP in regards to this study is that it collects very detailed information on personality traits and behavioral responses through a

[20] As in any longitudinal survey, attrition can an issue. By wave 2, 92% of the sample remained; by wave 3, 91% did so; by wave 4, 90%; and by wave 5, 86% remained in the sample. However, only the first three waves were used for most of the estimations presented in this paper. Appendix A presents an analysis on the attrited observations. In particular, being a bully or being a victim of bullies is not a determinant for leaving the sample.

Table 1: Descriptive Statistics

| | | | | *Incidence of Bullying* | |
|---|---|---|---|---|---|
| Total sample size | 3,449 | | | | |
| Number of Females | 1,724 | *Fathers Education:* | | Wave | |
| Urban households | 78.55% | High-school | 42.94% | 1 | .22499 |
| Single-headed hhs | 6% | 4yr Coll. or above | 36.56% | 2 | .11198 |
| Income | 1mill won | *Mothers Education:* | | 3 | .04768 |
| Students in tutoring | 81.82% | High-school | 56.31% | | |
| Single-child hhs | 8.6% | 4yr Coll. or above | 19.51% | | |

Note: Author's tabulations using KYP-JHSP. Income figure shown corresponds to the median monthly per-capita household income.

comprehensive battery of personality questions consistent across waves. The KYP-JHSP inquires about academic performance, student effort and participation in different kinds of private tutoring. The survey also asks about time allocation, leisure activities, social relations, attachment to friends and family, participation in deviant activity, and victimization in different settings including bullying. While the youths are often asked about the involvement of their parents in many aspects of their life, parents and guardians answer only a short questionnaire covering household composition and their education, occupation and income.

Bullying, as all other personal characteristic that were collected in the KYP-JHSP, is self-reported by the students. It refers to events where they have been severely teased or bantered, threatened, collectively harassed, severely beaten, or robbed. Hence, even though psychologists have constructed a very wide definition for bullying—presented in the introduction of this paper—the kids in the study respond to the most direct and less subtle versions of bullying.[21] This is in line with the findings in several international studies where children have been found to "focus on the more obvious and less subtle forms of bullying such as direct verbal and physical abuse and overlook indirect aggression" (Naylor et al., 2010).[22] In the same way, the reported incidence of bullying in the KYP-JHSP, presented in Table 1, is in line with other nationally representative studies (Kim et al., 2004), and with

[21]Nonetheless, under this limited definition, I find that there is at least one bully and one victim in every sampled classroom. This goes in line with the findings of Schuster (1999) in German schools.

[22]See, for instance, Madsen (1996); Smith and Levan (1995); Smith et al. (1999, 2002).

the incidence reported in international studies (see Smith and Brain (2000) for a summary). Furthermore, it closely mirrors the victimization incidence found in the US by the School Crime Supplement of the National Crime Victimization Survey (National Center for Education Statistics, 2015).

Data and institutional requirements aside, it is worth noting that—like in the US and many other countries in the world—bullying is a very important issue in the South Korean society, usually characterized by ultra-competitive academic environments that praise scholastic achievement.[23] Not surprisingly, such environments foster unhappiness and aggressiveness in the classrooms; a fertile ground for bullying. Given the link found between bullying and suicides (Kim and Leventhal, 2008; Kim et al., 2009), and the striking suicide rate among young people in South Korea,[24] the government has deployed active policies aimed at curving these phenomena.[25]

## 5.1 The Construction of the Manifest Measures for Identification of Unobserved Heterogeneity

As mentioned above while describing the empirical strategy, the estimation of the distribution parameters of the latent heterogeneity requires at least three manifest scores per factor. In this subsection, I present how those scores were constructed for each dimension of the unobserved heterogeneity.

[23] https://www.nytimes.com/2014/08/02/opinion/sunday/south-koreas-education-system-hurts-students.html?_r=0

[24] Suicide is the largest cause of death for people between 15 and 24, killing 13 for every 100,000 people in this age range. One school-aged kid (10 to 19 years old) commits suicide each day (Statistics Korea, 2012). Overall, South Korea has the single highest suicide rate in the world: 32 deaths per 100,000 people, according to the World Health Organization (http://www.who.int/gho/mental_health/suicide_rates_crude/en/).

[25] See http://www.bbc.com/news/world-asia-26080052. Reports indicate that since 2012, the government installed more than 100,000 closed-circuit cameras in school facilities to prevent bullying and prosecute its perpetrators.

### 5.1.1 Non-Cognitive Scores

To identify non-cognitive skills, I use measures of locus of control, responsibility and self-esteem. Most of the socio-emotional information in the KYP-JHSP is recorded in categories that group the reactions of the respondent in bins like "strongly agree" or "disagree". In consequence, and following common practice in the literature, I construct the socio-emotional manifest measures by adding the categorical answers across questions on the same topic.[26] This method incorporates some degree of smoothness in the scores, which is essential for the estimation procedure.

### 5.1.2 Cognitive Scores

The KYP-JHSP contains information on grades and academic performance. In particular, I use the grades obtained in tests of i) math and science; ii) language (Korean) and social studies; and iii) the score obtained in an overall test taken yearly. Previous literature has shown that grades may not be orthogonal to non-cognitive skills (Heckman et al., 2011). The production function of grades must be modeled using both cognitive and non-cognitive skills as inputs. As shown in Section 4, my model takes fully into account this feature of the data and incorporates it into the estimation.

### 5.1.3 The Construction of Measures on Skill Investment

I use measures of good parenting as indicator scores for investment choices in non-cognitive skills, namely parental physical and verbal abuse, parental control and

[26] To create the *locus of control* measure, I aggregated the answers to three questions: 1. I have confidence in my own decision; 2. I believe that I can deal with my problems by myself; 3. I am taking full responsibility of my own life. To create the *self-esteem* index I aggregated the answers to: 1. I think that I have a good character; 2. I think that I am a competent person; 3. I think that I am a worthy person; 4. Sometimes I think that I am a worthless person (the negative of); 5. Sometimes I think that I am a bad person (the negative of); 6. I generally feel that I am a failure in life (the negative of); 7. If I do something wrong, people around me will blame me much (the negative of); 8. If I do something wrong, I will be put to shame by people around me (the negative of). Finally, I created the *irresponsibility* index by adding the answers to the following questions: 1. I jump into exciting things even if I have to take an examination tomorrow; 2. I abandon a task once it becomes hard and laborious to do; 3. I am apt to enjoy risky activities.

parental harmony. The first measure indicates how often is the child beaten, physically hurt, yelled at or addressed in an inappropriate manner by the parents. Parental control relates to how well parents know where the kid is, who is she with, what is she doing and when is she coming back home. Parental harmony collects information related the level of care and interest in her life the kid feels from her parents.[27]

The measures used to identify the cognitive skill investment factor relate to the enrollment in private tutoring of each kid. South Korean society is characterized by the high importance it gives to academic success. Hence, it is not uncommon for kids to be enrolled in after-school academic programs. By age 14, around four fifths of the sample attend some kind of tutoring. As manifest variables of cognitive skill investment I use a scale of how personalized tutoring session are,[28] the time spent in tutoring, and the cost of the tutoring.

# 6 Results[29]

As explained in Section 4.2, estimation was divided into two stages.[30] For that reason, I used the Limited Information Maximum Likelihood (LIML) technique to correct the standard errors of the second stage (Greene, 2000). Common controls to all the equations in the structural model were: age, gender, family composition—number of older and younger siblings, urban status, broken home status, father's education—and

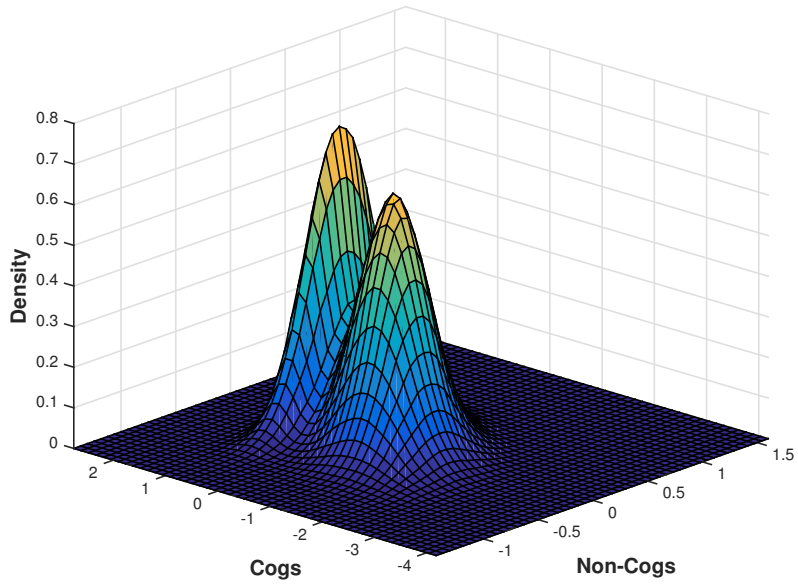
[27]See Appendix C for a detailed explanation of the questions used to create each score

[28]This manifest score collects information of the nature of the extra-school classes taken. That is, whether the classes were entirely private, with few classmates, with many classmates, or through the internet. Students gave this type of information about their tutoring for every subject (e.g., language, math, science), and based on that I created aggregated measures.

[29]To keep the paper within a reasonable length, I placed some of the background estimates and tables with the complete set of controls in the Web Appendix available at https://goo.gl/G56a9u.

[30]First stage estimations show that skill distributions for $t$ and $t+1$ are far from normal, and that there is a positive correlation between both dimensions of skills: 0.4499 and 0.358, respectively. This indicates that kids with high levels of one skill tend to have high levels of the other skills as well. Interestingly, I find that the variance of non-cognitive skills increases for higher levels of cognitive-skills. Hence, socio-emotional abilities, although positively correlated with cognitive skills, are less so for smarter kids. Full set of parameter estimates can be found in Table 1 and Table 2 and Figures 1(a) and 1(b) in Section 2.1 of the Web Appendix.

Figure 2: Initial Skills Distribution




Note: Distributions constructed from parameters that resulted from the estimation of measurement system (15). Complete set of estimates can be found in Tables the Web Appendix. Factor distributions estimated using a mixture of two normals. The estimated parameters of those normals are: $\sigma_{1,NC} = 0.272$, $\sigma_{2,NC} = 0.630$, $\sigma_{1,C} = 0.321$, $\sigma_{2,C} = 0.217$, $\mu_{1,NC} - 0.076$, $\mu_{1,C} = -0.529$, $corr\left(\theta_{NC}, \theta_C\right)_1 = 0.749$, $corr\left(\theta_{NC}, \theta_C\right)_2 = 0.597$ and $p = 0.647$, where $p$ is the mixing probability.

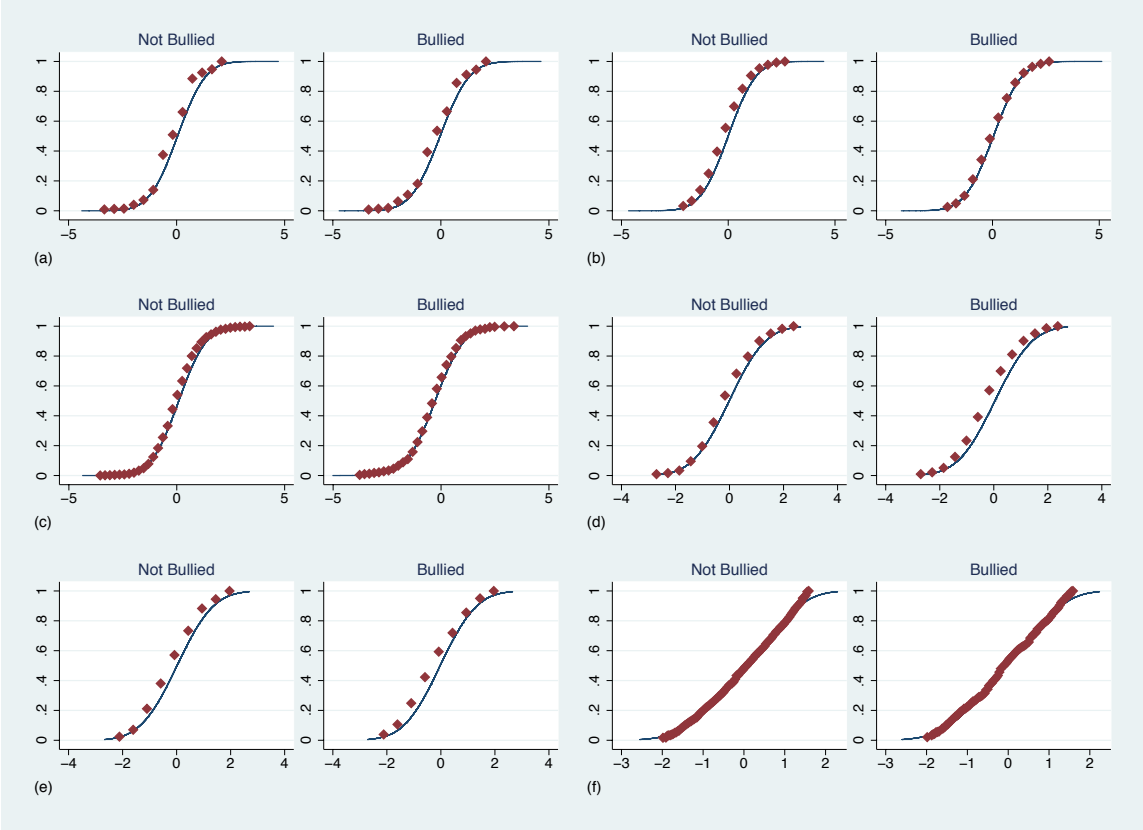
per-capita household monthly income. Additional variables, specific to each equation (i.e., exclusion restrictions) will be explained below.

## 6.1 Model Fit

In the first step, I estimate the initial distribution of skills from model (15) which incorporates the structure proposed by Hansen et al. (2004) to address the possible problems of joint causality. Figure (2) presents the estimated initial distribution of skills. As expected cognitive and non-cognitive skills are positively correlated: $corr\left(\theta_{NC,t}, \theta_{C,t}\right) = 0.450$.

Table 2 and Figures 3 show that the model fits extremely well the actual data. The former not only shows that the incidence of bullying is matched almost exactly,

Figure 3: Actual vs. predicted test scores cumulative distributions conditional on victimization at $t = 1$




Note: Actual (diamond) and predicted (line) cumulative distributions plotted of the following test scores: (a) locus of control (b) irresponsibility (c) self esteem (d) language and social studies (e) math and sciences (f) year exam. The predicted values come from simulations based on the estimated parameters of the model

but also that the means and standard errors of the simulated test scores are very close to the ones obtained from the actual scores, for each victimization state (i.e., bullied or not bullied). In fact, I cannot reject the null of equality of means in any of the 12 cases.

Figures 3 plot the predicted test scores provided by the model against the actual CDF of each test observed in the data. The figures show a remarkable fit in all scores, regardless of the victimization condition. I corroborate this by performing a Kolmorov-Smirnov test on the predicted and actual score distributions. The results

Table 2: Goodness-of-fit of the model

| | Wave1 | Wave2 |
|---|---|---|
| *Bullying* | | |
| Actual | 0.2267 | 0.1112 |
| Predicted | 0.2264 | 0.1159 |

| | Locus | Irresp. | SelfEst | Lang. | Math | YrScr |
|---|---|---|---|---|---|---|
| *Not Bullied Students* | | | | | | |
| *Means* | | | | | | |
| Actual | 0.0182 | -0.0491 | 0.0986 | 0.0411 | 0.0258 | 0.0315 |
| Predicted | 0.0433 | -0.0272 | 0.0954 | 0.0084 | 0.0148 | 0.0134 |
| *Std. Devs.* | | | | | | |
| Actual | 0.9847 | 0.9860 | 0.9704 | 0.9905 | 0.9798 | 0.9972 |
| Predicted | 1.0202 | 1.0211 | 1.0056 | 1.0457 | 1.0537 | 1.0431 |
| *K-S p-value* | 0.1329 | 0.2179 | 0.0080 | 0.5825 | 0.8008 | 0.3017 |
| *Bullied Students* | | | | | | |
| *Means* | | | | | | |
| Actual | -0.0518 | 0.1483 | -0.2698 | -0.0470 | -0.0184 | -0.0805 |
| Predicted | -0.0349 | 0.1191 | -0.2719 | -0.0313 | -0.0233 | -0.0801 |
| *Std. Devs.* | | | | | | |
| Actual | 1.0689 | 1.0373 | 1.0497 | 1.0158 | 1.0508 | 0.9948 |
| Predicted | 1.0642 | 1.0491 | 1.0510 | 1.0679 | 1.0535 | 1.0463 |
| *K-S p-value* | 0.2684 | 0.4537 | 0.0142 | 0.6274 | 0.9258 | 0.5650 |

Note: Predicted means are not statistically different from the actual means at any conventional level of significance. `Locus` stands for locus of control score. `Irresp`. stand for irresponsibility score. `SelfEst` stands for self esteem score. `Lang` stands for language and social studies score. `Math` stands for math and sciences score. `YrScr` stand for the year exam score. The predicted values come from simulations based on the estimated parameters of the model

Table 3: Dynamic Estimation

| | (1) | (2) | (3) | | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|---|
| | | | | | $M_{t+1} = 0$ | | $M_{t+1} = 1$ | |
| | $M_{t+1}$ | $I_{NC,t+1}$ | $I_{C,t+1}$ | | $\theta_{NC,t+1}$ | $\theta_{C,t+1}$ | $\theta_{NC,t+1}$ | $\theta_{C,t+1}$ |
| $\theta_{NC,t}$ | -0.071** | 0.675*** | 0.108 | $\theta_{NC,t}$ | 0.952 | 0.105 | 0.904 | 0.087 |
| | (0.028) | (0.075) | (0.110) | | (0.016) | (0.016) | (0.041) | (0.041) |
| $\theta_{C,t}$ | 0.008 | 0.028 | 0.343*** | $I_{t+1}$ | 0.031 | 0.057 | 0.030 | 0.049 |
| | (0.009) | (0.022) | (0.036) | | (0.012) | (0.008) | (0.029) | (0.023) |
| $\nabla\left(\hat{\theta}_{NC,t}\right)$ | -0.007*** | | | $\rho$ | -0.084 | -0.032 | 0.357 | -0.196 |
| | (0.003) | | | | (0.084) | (0.042) | (0.347) | (0.145) |
| $\nabla\left(\hat{\theta}_{C,t}\right)$ | 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 (7) of the structural model. See complete estimates in Table 5 in Section 2.3 of the Web Appendix. It includes region fixed-effects and 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 as described in footnote 17). $\nabla(\cdot)$ refers to the number of classmates within a window of 10% of a SD around observation $i$. $\hat{\theta}_{NC}$ is the residualized measure of self-esteem and $\hat{\theta}_C$ 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 (5) and (6) of the structural model. Columns (4)-(7) present the estimates of equations (1), for victimization-specific production function of non-cognitive and cognitive skills. Note that the coefficient for $\theta_{C,t}$ (i.e., $\gamma_{C,t}$) can be obtained from $\gamma_{C,t} = 1 - \gamma_{NC,t} - \gamma_{I,t}$.

found in Table 2 indicate that the predicted scores come from a distribution that is not different from the one the actual scores describe in ten out of the twelve comparisons. The only score for which I fail the K-S test is self-esteem, even though the first and second moments are closely matched. I suspect this is the case because the actual self-esteem distribution has a kink or jump close to the median which is difficult to fit with smooth and continuous latent factors. In fact, when the observed distributions of self-esteem are smoothed using a kernel approximations, the K-S test statistics get closer to the non-rejection values.

## 6.2 Results from the Dynamic Model

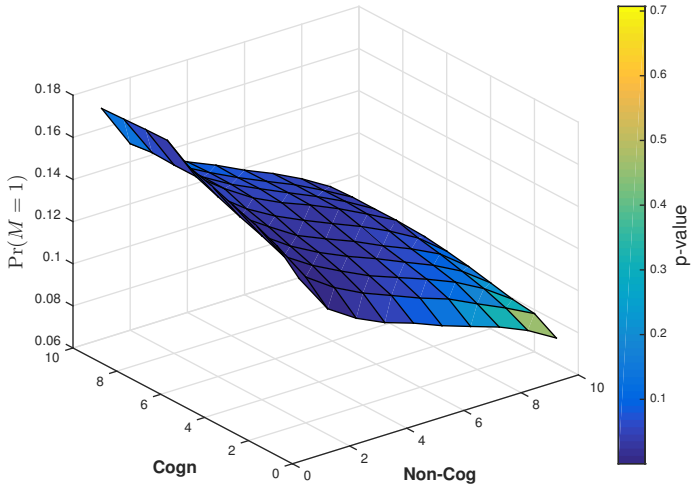
### 6.2.1 Incidence of victimization

Column 1 in Table 3 shows the relation between skills and selection into bullying. In line with the results of Sarzosa and Urzua (2015), kids with less non-cognitive skills are significantly more likely to be bullied. A one standard deviation decrease in non-cognitive skills increases the likelihood of being victimized by 2.26 percentage points. It represents an increase in the probability of being victimized of about a fifth. Column 1 in Table 3 also shows the importance the relation between own and peer characteristics has in determining peer victimization. Controlling for observable characteristics and skill levels, kids who were placed in a school in which their non-cognitive skills are uncommon are significantly more likely to be bullied. The results indicate that for each additional classmate with similar non-cognitive skill endowments she has the average student's likelihood of victimization drops by one percentage point. Interestingly, uncommonness in terms of income also encourages victimization. Bullying probability falls by half a percentage point for each additional classmate that has a family income level similar to the one of the prospective victim. These results are remarkably robust to the inclusion of the percentage of classmates that come from troubled families and the percentage of bullies in the classroom.[31]

The fact that the model relies on the identification of unobserved heterogeneity allows me to quantify the victimization probability no only for the average student, but also for every combination of skills at a given point in time. Figure 4 shows striking differences in the likelihood of being bullied depending on the level of non-cognitive skills. Kids in the first decile of non-cognitive skills are twice more likely to be bullied than those in the tenth decile, and are 36% more likely to be bullied that the average student. In addition, Figure 4 shows that among those with low non-cognitive skills, the ones that have higher cognitive skills are 3 percentage points more likely to be victimized than those in the bottom of the cognitive skill distribution.

[31] See the robustness checks in Section 2.3 in the Web Appendix.

Figure 4: Probability of Being Bullied




Note: Results based on 40,000 simulations based on the estimated parameters of the dynamic model.

This reflects the widely held notion that socially-awkward-smart children face greater chances of being victimized in school.

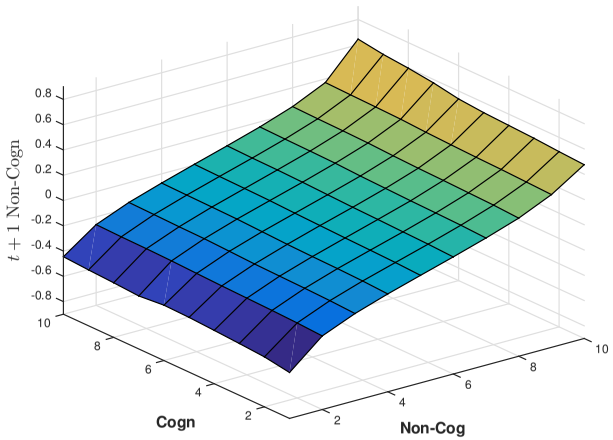
### 6.2.2 Skills Production

Columns 4 to 7 in Table 3 present the results of estimating the system described by (1). They contain the parameters—that together with the ones related to selection into bullying and the distributions of the unobserved heterogeneity—govern the dynamic process of skill formation.[32] This process is presented in Figures 5. Figures 5a and 5b show that high non-cognitive skills produce high future non-cognitive skills, and that marginal increments of those initial skills are very productive (i.e., non-cognitive skills self-productivity $\partial\theta_{t+1}^{NC}/\partial\theta_t^{NC} > 0$ for the entire $\left(\theta_t^{NC}, \theta_t^{C}\right)$ space). These figures also demonstrate that cognitive skills are unimportant in the non-cognitive skill production process except for the fact that higher initial cognitive skills make the marginal increments of the initial non-cognitive skills more productive (i.e., $\partial^2\theta_{t+1}^{NC}/\partial\theta_t^{NC}\partial\theta_t^{C} > 0$).

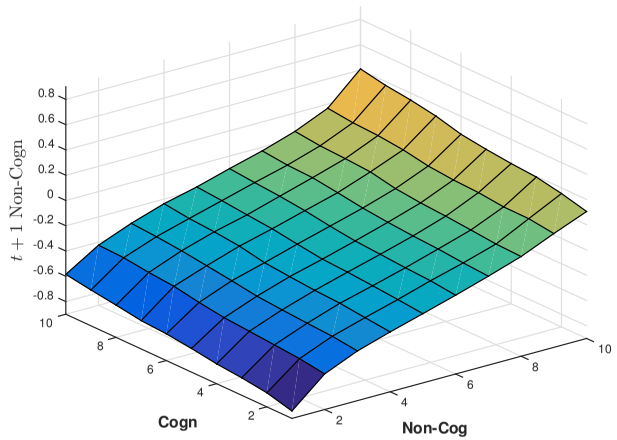
[32]Details and results of the estimation of the latent factors of parental investments can be found in Section 2.2 in the Web Appendix.

Figure 5: $\theta^S_{t+1}$ as a function of $\theta^{NC}_t$ and $\theta^C_t$

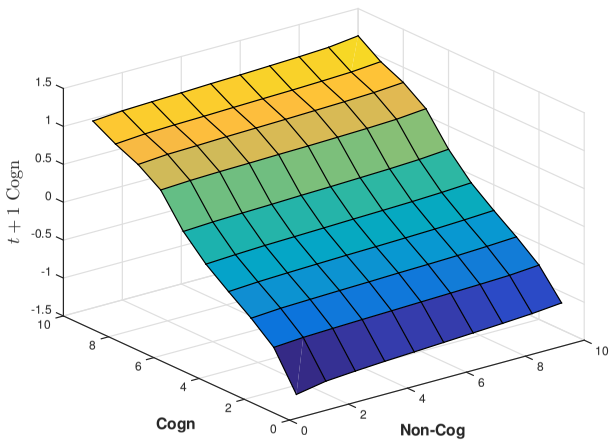
(a) Non-Cognitive: $M_{t+1} = 0$



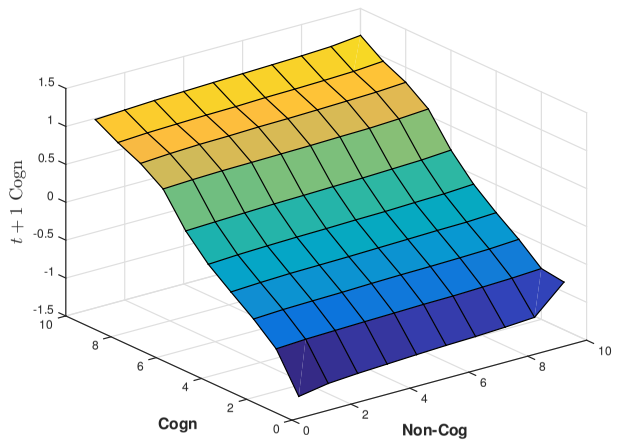

(b) Non-Cognitive: $M_{t+1} = 1$




(c) Cognitive: $M_{t+1} = 0$




(d) Cognitive: $M_{t+1} = 1$




Note: Results based on 40,000 simulations based on the estimated parameters of the dynamic model.

Table 4: ATE of Being Bullied on Next Period Skills for the Average Student

| $\theta_{NC,t+1}$ | | $\theta_{C,t+1}$ | |
|---|---|---|---|
| Estimated | -0.249*** | Estimated | -0.009 |
| | (0.020) | | (0.019) |
| As StdDev of $\theta_{NC,t+1}$ | -0.399 | As StdDev of $\theta_{C,t+1}$ | -0.007 |

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The Table estimates $E\left[\theta_{t+1}^{NC}|\bar{\theta}_t^{NC}, \bar{\theta}_t^{N}, M_{t+1}=1\right] - E\left[\theta_{t+1}^{NC}|\bar{\theta}_t^{NC}, \bar{\theta}_t^{N}, M_{t+1}=0\right]$ using 40,000 simulations based on the estimated parameters of the dynamic model. Standard deviation of $\theta_{t+1}^{NC} = 0.623$.

Figures 5c and 5d show that the production of cognitive skills relies heavily on past levels of cognitive skills. Although the existing levels of non-cognitive skills contribute to the production process of cognitive skills, their contribution is small compared to that of the existing stock of cognitive skills. For instance, going from decile 1 to decile 10 in non-cognitive skills distribution has the same effect on the production of cognitive skills as increasing the cognitive skills input by one decile.

My results indicate that there is a strong path dependence in which skills produce skills, setting a high cost in terms of future stock of skills for those who start the accumulation process in the lower quantiles of the skill distribution. My results also show that this path dependence is not reversed by investment choices. In fact, Columns 2 and 3 in Table 3 show that investment choices in non-cognitive skills depend greatly on the past level of non-cognitive skills, and investment choices in cognitive skills depend greatly on past levels of that skill in the first place. Hence, people with high skills not only pass their high stock on to the next period, but also they are more prone to invest in their development.

### 6.2.3 Effects of Bullying on Skill Production

Table 4 shows the effect of bullying on the accumulation of cognitive and non-cognitive skills. To calculate this, I compare the next period skills of those who would be

selected into bullying with those who would not, evaluated at the skills' means. That is, $E\left[\hat{\theta}^S_{t+1}|\bar{\theta}^{NC}_t, \bar{\theta}^N_t, M_{t+1}=1\right] - E\left[\hat{\theta}^S_{t+1}|\bar{\theta}^{NC}_t, \bar{\theta}^N_t, M_{t+1}=0\right]$ for $S = \{NC, C\}$. I find that bullying impedes non-cognitive skills accumulation by -0.249. That is equivalent to a reduction in non-cognitive skill accumulation of 39.9% of a standard deviation. This is a sizable effect. It implies a reduction of 33.6% of a standard deviation in the language test score, and a reduction of 28.9% of a standard deviation in the math test score. These skill loses imply that the average kid would be 19 percentage points more likely to report being sick recently, 5.5 percentage points more likely to smoke and 10.5 more likely to drink alcoholic beverages. The stock of skills lost also translates to setbacks in mental health. They equate to an increase of 48.77% of a standard deviation in the depression symptom scale, an increase of 38.1% of a standard deviation in the levels of stress cased by insecurities regarding his or her image, and a third of a standard deviation in the levels of stress caused by issues regarding school.[33]

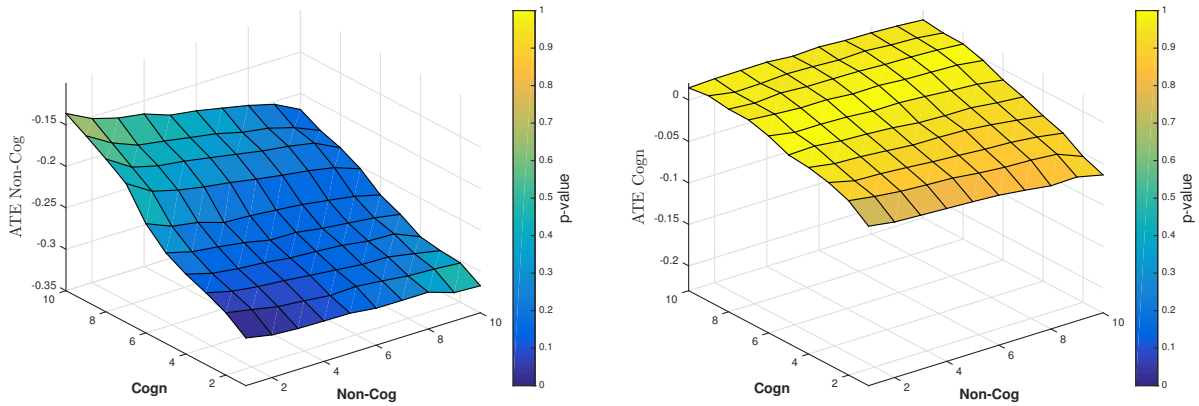
The same estimation shows there is no statistically significant effect of bullying on cognitive skill accumulation. These results indicate that, as expected, bullying is much more costly in the non-cognitive dimension than in the cognitive one. Although victims might skip school, their learning ability is not affected as gravely as their ability to self regulate, overcome obstacles, see themselves positively or relate with others. Note that even if cognitive skills are unaffected, grades drop because of the effect non-cognitive skills have on them.

**Analyzing the effect beyond the mean.** Figure 6a presents the effect of bulling in next period non-cognitive skills for each level of initial skills. It shows that the kids that suffer the greatest negative impact are those who start the dynamic process with low stocks of skills. While victims with low levels of skills loose almost half of

[33]In Section 2.4 of the Web Appendix, I present detailed result of estimating models of unobserved heterogeneity at age 16 of the form $Y = \mathbf{X}_Y\beta^Y + \alpha^{Y,NC}\theta^{NC} + \alpha^{Y,C}\theta^C + e^Y$, where $Y$ is depression, stress in different situations, and the likelihood of smoking, drinking alcohol, felling healthy, being satisfied with life, or going to college by age 19.

Figure 6: $E\left[\theta^S_{t+1}|\theta^{NC}_t, \theta^N_t, M_{t+1}=1\right] - E\left[\theta^S_{t+1}|\theta^{NC}_t, \theta^N_t, M_{t+1}=0\right]$

(a) Non-Cognitive (b) Cognitive




Note: Results based on 40,000 simulations based on the estimated parameters of the dynamic model.

a standard deviation of non-cognitive skills, victims with high stocks of skills loose a third of a standard deviation. In particular, those who start with low cognitive skills face harsher consequences. However, due to the positive correlation between cognitive and non-cognitive skills, those with low cognitive skills are very likely to also have low levels of non-cognitive ones. Such treatment effect heterogeneity based on the initial levels of skills, and the fact that victimization also depends on them yield a very interesting result: kids with low initial levels of skills are not only more likely to be bullied, but also its consequences are stronger on them. Table 5 attest to that. It shows how bullying shifts students to lower deciles of the next-period skills distribution. In fact, it shows that if students from the lower 40% of the non-cognitive skill distribution at $t$ were to be victimized, they would end up belonging to the lowest non-cognitive skills decile at $t+1$. Furthermore, if students from the bottom 80% of the non-cognitive skill distribution at $t$ were to be victimized, they would end up belonging to the lowest half of the non-cognitive skills distribution in the next period. Notably, those that start with abundant stocks of skills fall less far from their original place in the skills distribution. Victims from the top decile at $t$ end up in the ninth

Table 5: Decile of $\theta_{t+1}^{NC}$ that students would end up if victimized in $t+1$, by initial skills decile

| $Q_{10}(\theta_t^{NC})$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Pr(M=1)$ | 0.153 | 0.136 | 0.128 | 0.122 | 0.117 | 0.112 | 0.107 | 0.102 | 0.096 | 0.085 |
| $E[\theta_{NC,t+1}\|M=1]$ | -0.751 | -0.542 | -0.433 | -0.346 | -0.269 | -0.189 | -0.108 | -0.014 | 0.109 | 0.344 |
| $Q_{10}(\theta_{t+1}^{NC})$ | 1 | 1 | 1 | 1 | 2 | 3 | 4 | 5 | 7 | 9 |

Note: $Q_{10}(x)$ stands for decile of $x$. Estimations obtained from 40,000 simulations based on the parameter estimates of the dynamic model.

decile at $t+1$.

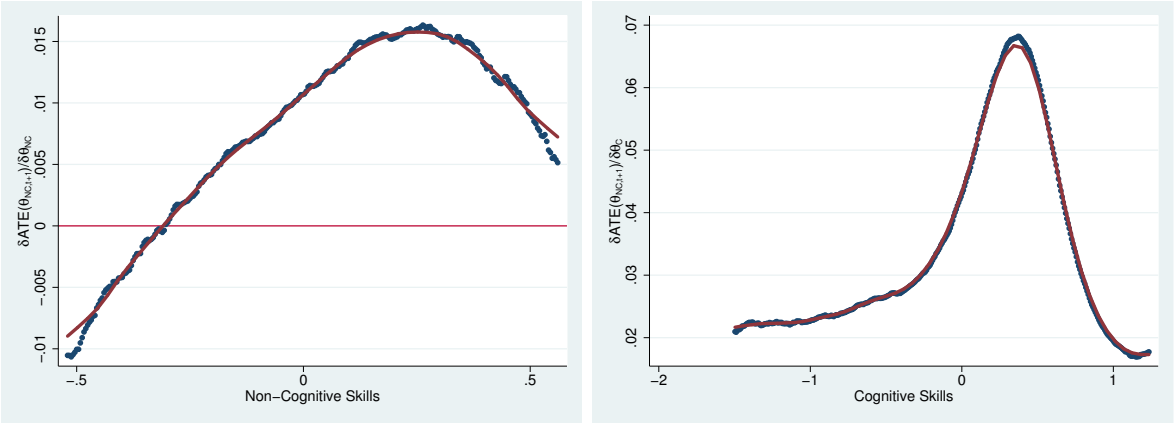
### 6.2.4 Complementarities

As explained in Section 3, an important feature of the model is that it allows the analysis of complementarities between skills and bullying. Namely, the measurement of how much the effect of bullying is modified by a marginal change in previous period skills (i.e., $\partial\left(\frac{\Delta\theta_{S,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{S',t}$ for $S=\{NC,C\}$ and $S'=\{NC,C\}$). Figures 7 show that marginally increasing initial levels of non-cognitive skills will result in small reductions in the negative effect of bullying on future period skills. This attest to the fact that the impact of bullying is relatively constant across the entire non-cognitive skills distribution. Figure 7a also shows that the palliation of the negative effect due to a marginal increase in non-cognitive skills is larger for those with above the mean initial non-cognitive skills.

On the other hand, Figure 7b shows that marginal increases in initial levels of cognitive skills have larger effects in palliating the negative effect of bullying on non-cognitive skills. In fact, it would be reduced by four percentage points or 16% for the average kid. For those in the sixth and seventh decile, the palliation effect is even larger, reducing the negative effect of bullying on non-cognitive skills by about a fourth.

Figure 7: Static Complementarity

(a) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$ (b) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{C,t}$




Note: Results based on 40,000 simulations based on the estimated parameters of the dynamic model. The scatter plot presents the static complementarity measures at 750 points along the skill distributions. The line represents a local polynomial approximation.

All the evidence presented in this paper argues in favor of the existence of a self-reinforcing mechanism in which-low skilled kids are more likely to be victims of violence at their schools, and in turn not only their skills are depleted by the bullying event itself, but also its consequences aggravated for those who started with low skill levels in the first place. This sends them in a downward spiral by making them even more at risk of being victims of bullying in the future, which in turn will be much more harmful events, and therefore having always more and more difficulties in acquiring the non-cognitive skills they lack. Even though, I show that investment in skills during middle school years is often unproductive, the static complementarity results suggest that even a tiny bit of skill accumulation would have an immense impact not only in deterring bullying, but also in lessening its consequences among those that are more at risk.

# 7 Policy Implications

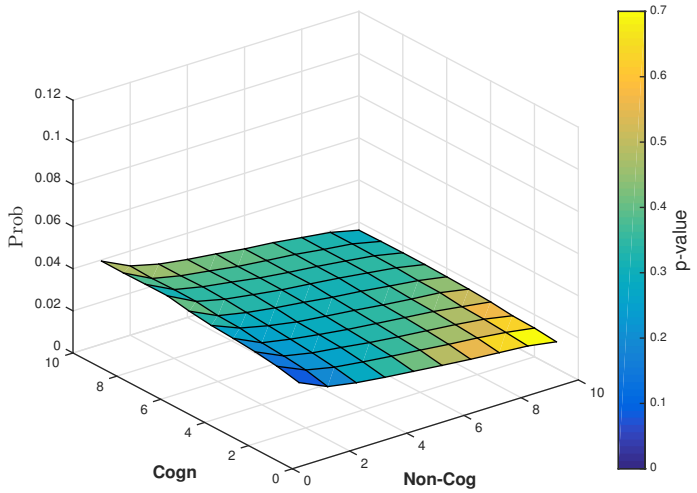
Several anti-bullying campaigns have been deployed all around the world in an ambitious effort to eliminate this unwanted phenomenon.[34] My findings indicate there are at least two fronts policymakers can work on. First, the development of non-cognitive skills. Non-cognitive skilled kids will be more less likely to be victimized. And if they happen to be bullied, its impact on their skill accumulation path is much lessened. The importance of developing non-cognitive skills at young ages is heightened by the strong dependance of current skill levels on past skills levels.

The second implication of my results relates to classroom assignment. Column 1 in Table 3 shows that, given skill levels and observable characteristics, children with uncommon traits are more likely to be targeted by bullies. This leads to policy-relevant question: to what extent can allocating children to more homogenous classrooms deter victimization? To answer this, I simulate the model with an extreme—unfeasible in practice—mechanism of allocating students to classrooms, Consider it a benchmark scenario. It places students in classrooms with kids that have similar stocks of non-cognitive skills, as measured by the self-esteem score. This exercise ignores geographical distances. It just sorts the universe of students with respect to their self-esteem scores and split them in classrooms according to the typical classroom size in South Korea.

Figure 8 presents the results of these simulations. As in Figure 4, it plots the likelihood of being bullied for every skill level. A comparison between these two figures shows the massive impact that reducing in-classroom non-cognitive skill heterogeneity has on the likelihood of being victimized. The benchmark case in Figure 8 shows that by arranging students with classmates that have similar levels of non-cognitive skills, the overall likelihood of victimization falls from 11.5% to 2.8%. This dramatic reduction is across the entire skills domain to the point that almost everyone has

[34] See the Olweus Bullying Prevention Program and the US Education Department stopbullying.gov program.

Figure 8: Classroom Allocation Simulations: Benchmark




Note: Results based on 40,000 simulations based on the estimated parameters of the dynamic model.

a probability of being victimized that is not statistically different form zero. Only those who start the period with very low levels of non-cognitive skills would still face a non-zero likelihood of being bullied of around 4%. However, they would face a sizable reduction in their hazard of being bullied in the order of 11 percentage points.

# 8 Conclusions

This paper develops and estimates a structural model of skill accumulation that introduces endogenous social interactions as a driver of the dynamic process. The model uses several dimensions of unobserved heterogeneity and in-classroom variation of student characteristics to identify the endogenous selection of bullying victims. My findings indicate the existence of a vicious cycle between victimization and skill depletion. I find that bullying is disproportionately suffered by students that lack socio-emotional skills, and among those, the smart students are more likely to be victimized. My findings, in line with psychological studies, suggest that conditional

on the level of skills, kids with uncommon characteristics relative to those of their classmates are more likely to be victimized.

The dynamic estimation showed that bullying is very costly in terms of the amount of skills lost from one period to the next. Bulling at age 15 reduces non-cognitive skill accumulation by a 40% of a standard deviation for the average kid. That effect is a third greater for kids with low initial levels of skills. Static complementarity shows current stock of cognitive skills influences greatly the "negative productivity" of the bullying event.

These results show the existence of a self-reinforcing mechanism, in which initial levels of skill become crucial, suggesting that policies aimed to foster cognitive skills at early ages will greatly reduce victimization occurrence. In addition, my model indicates that the allocation of students in more homogeneous classroom might reduce victimization by preventing kids with uncommon characteristics to be isolated and targeted by bullies.

This paper intends to contribute to the human development literature in economics by exploring how victimization of school-aged kids may hamper the development of successful adults. In the process, this paper contributes to the skill formation literature by introducing endogenous social interactions as triggers of phenomena that have long-lasting consequences. This paper opens a promising research agenda. For instance, researchers can use the model to analyze other types of disruptive behaviors, the role that gender plays in classroom dynamics *vis-a-vis* those social interactions, or—data permitting—the introduction of physical traits as determinants of victimization. Furthermore, given the importance of initial levels of skills, we should inquire about how these negative social interactions affect skill accumulation among younger children.

# References


Agostinelli, F. and Wiswall, M. (2016a). Estimating the Technology of Children's Skill Formation. *NBER Working Paper*, (22442).

Agostinelli, F. and Wiswall, M. (2016b). Identification of Dynamic Latent Factor Models.

Aizer, A. and Cunha, F. (2012). The Production of Human Capital: Endowments, Investments and Fertility. *NBER Working Paper*, (18429).

Angrist, J. (2014). The perils of peer effects. *NBER Working Paper*, (19774).

APA (2006). The american psychological association dictionary. *American Psychological Association Reference Books.*

Attanasio, O., Meghir, C., and Nix, E. (2017). Human Capital Development and Parental Investment in India.

Baron, S. (2016). Play Therapy with Spectrum of Bullying Behavior. In Cramershaw and Stewart, editors, *Play Therapy.* New York.

Bartholomew, D., Knott, M., and Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach. *Wiley Series in Probability and Statistics. West Sussex, United Kingdom.*

Björkqvist, K., Ekman, K., and Lagerspetz, K. (1982). Bullies and victims: Their ego picture, ideal ego picture and normative ego picture. *Scandinavian Journal of Psychology*, 23(1):307–313.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Weel, B. T. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4):972–1059.

Brown, S. and Taylor, K. (2008). Bullying, education and earnings: evidence from the national child development study. *Economics of Education Review*, 27(4):387–401.

Cabrera, N., Shannon, J., and Tamis-LeMonda, C. (2007). Fathers' influence on their children's cognitive and emotional development: From toddlers to pre-k. *Applied Development Science*, 11(4):208–213.

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

Carrell, S. E. and Hoekstra, M. (2012). Family Business or Social Problem? The Cost of Unreported Domestic Violence. *Journal of policy analysis and management*, 31(4):861–875.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids. *American Economic Journal: Applied Economics*, 2(1):211–228.

Carrell, S. E., Hoekstra, M. L., and Kuka, E. (2016). The Long-Run Effects of Disruptive Peers.

Cawley, J., Heckman, J., and Vytlacil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4):419–442.

Cunha, F., Heckman, J., Lochner, L., and Masterov, D. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Dake, J., Price, J., and Telljohann, S. (2003). The nature and extent of bullying at school. *Journal of school health*, 73(5):173–180.

Doyle, O., Harmon, C., Heckman, J., and Tremblay, R. (2009). Investing in early human development: timing and economic efficiency. *Economics & Human Biology*, 7(1):1–6.

Duckworth, A. and Seligman, M. (2005). Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological Science*, 16(12):939–944.

Eriksen, T. L. M., Nielsen, H. S., and Simonsen, M. (2014). Bullying in elementary school. *Journal of Human Resources*, 49(4):839–71.

Espinoza, R., Sarzosa, M., and Urzua, S. (2014). The causal effect of skills: An international study. *University of Maryland*, pages 1–53.

Faris, R. and Felmlee, D. (2011). Status struggles network centrality and gender segregation in same-and cross-gender aggression. *American Sociological Review*, 76(1):48–73.

Flouri, E. and Buchanan, A. (2002). Life satisfaction in teenage boys: The moderating role of father involvement and bullying. *Aggr. Behav.*, 28(2):126–133.

Fredriksson, P., Ockert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics*.

Greene, W. H. (2000). *Econometric analysis*. Prentice Hall, Upper Saddle River, New Jersey, 4th edition.

Gunnison, E., Bernat, F. P., and Goodstein, L. (2016). *Women, Crime, and Justice*. Balancing the Scales. John Wiley & Sons.

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.

Hart, B. and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*.

Heckman, J., Humphries, J., and Mader, N. (2011). The ged. *Handbook of Economics of Education*, 3.

Heckman, J. and Masterov, D. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3):446–493.

Heckman, J. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the ged testing program. *The American Economic Review*, 91(2):145–149.

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

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4):451–464.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, 94(1-2):114–128.

Hodges, E., Malone, M., and Perry, D. (1997). Individual risk and social risk as interacting determinants of victimization in the peer group. *Developmental psychology*, 33(6):1032.

Jackson, C. (2013). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina. *works.bepress.com*.

Judd, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, Massachusetts.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from south korea. *Journal of Urban Economics*, 61(3):458–495.

Kann, L., Kinchen, S., Shanklin, S. L., Flint, K. H., Hawkins, J., Harris, W. A., Lowry, R., OMalley Olsen, E., McManus, T., Chyen, D., Whittle, L., Taylor, E., Demissie, Z., Brener, N., Thornton, J., Moore, J., and Zaza, S. (2014). Youth Risk Behavior Surveillance—United States, 2013. 63(4).

Kiernan, K. E. and Huerta, M. C. (2008). Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood1. *The British Journal of Sociology*, 59(4):783–806.

Kim, Y. S., Koh, Y.-J., and Leventhal, B. L. (2004). Prevalence of School Bullying in Korean Middle School Students. *Archives of Pediatrics & Adolescent Medicine*, 158(8):737–741.

Kim, Y. S. and Leventhal, B. (2008). Bullying and suicide. A review. *International Journal of Adolescent Medicine and Health*, 20(2):133–154.

Kim, Y. S., Leventhal, B. L., Koh, Y.-J., and Boyce, W. T. (2009). Bullying Increased Suicide Risk: Prospective Study of Korean Adolescents. *Archives of Suicide Research*, 13(1):15–30.

Knudsen, E., Heckman, J., Cameron, J., and Shonkoff, J. (2006). Economic, neurobiological, and behavioral perspectives on building america's future workforce. *Proceedings of the National Academy of Sciences*, 103(27):10155–10162.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76.

Ladd, G. and Ladd, B. (1998). Parenting behaviors and parent–child relationships: Correlates of peer victimization in kindergarten? *Developmental psychology*, 34(6):1450.

Lowenstein, L. F. (1978). The bullied and non-bullied child. *Bulletin of the British Psychological Society*, (31):316–318.

Madsen, K. C. (1996). Differing perceptions of bullying and their practical implications. *Education and Child Psychology*, 13(2):14–22.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Manski, C. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115–136.

McEwen, B. and Seeman, T. (2006). Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 896(1):30–47.

Murnane, R., Willett, J., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, 77(2):251–266.

National Center for Education Statistics (2015). *Student Reports of Bullying and Cyber-Bullying: Results From the 2013 School Crime Supplement to the National Crime Victimization Survey*, volume 2015. US Department of Education.

Naylor, P., Cowie, H., Cossin, F., Bettencourt, R., and Lemme, F. (2010). Teachers' and pupils' definitions of bullying. *British Journal of Educational Psychology*, 76(3):553–576.

OECD (2014). Skills for social progress. pages 1–103.

Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education*, 12(4):495–510.

Payton, J., Weissberg, R., Durlak, J., Dymnicki, A., Taylor, R., Schellinger, K., and Pachan, M. (2008). The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews. technical report. *Collaborative for Academic, Social, and Emotional Learning (NJ1)*.

Pollak, R. (1976). Interdependent preferences. *The American Economic Review*, pages 309–320.

Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., and Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggr. Behav.*, 22(1):1–15.

Saltiel, F., Sarzosa, M., and Urzúa, S. (2017). Cognitive and socio-emotional abilities. In Jones, G., Johnes, J., Agasisti, T., and Lopez Torres, L., editors, *Handbook of Contemporary Education Economics*. Cheltenham, UK.

Sarzosa, M. and Urzua, S. (2015). Bullying in teenagers, the role of cognitive and non-cognitive skills. *NBER Working Paper*, w21631.

Schelling, T. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Schuster, B. (1999). Outsiders at school: The prevalence of bullying and its relation with social status. *Group Processes & Intergroup Relations*, 2(2):175–190.

Skinner, E. and Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, 85(4):571.

Smith, P. and Brain, P. (2000). Bullying in schools: Lessons from two decades of research. *Aggressive behavior*, 26(1):1–9.

Smith, P., Cowie, H., Olafsson, R., and Liefooghe, A. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen–country international comparison. *Child development*, 73(4):1119–1133.

Smith, P. and Levan, S. (1995). Perceptions and experiences of bullying in younger pupils. *British Journal of Educational Psychology*, 65(4):489–500.

Smith, P., Madsen, K., and Moody, J. (1999). What causes the age decline in reports of being bullied at school? towards a developmental analysis of risks of being bullied. *Educational Research*, 41(3):267–285.

Smith, P., Talamelli, L., Cowie, H., Naylor, P., and Chauhan, P. (2004). Profiles of non-victims, escaped victims, continuing victims and new victims of school bullying. *British Journal of Educational Psychology*, 74(4):565–581.

Statistics Korea (2012). *Annual Report on the Cause of Death Statistics*, volume 2011. Statistics Bureau, Economic Planning Board.

Tough, P. (2012). *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. Houghton Mifflin Harcourt.

Urzua, S. (2008). Racial labor market gaps. *Journal of Human Resources*, 43(4):919.

# 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.

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 | |

*** p<0.01, ** p<0.05, * p<0.1

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

| | Coef | StdErr | | Coef | StdErr | | Coef | StdErr |
|---|---|---|---|---|---|---|---|---|
| Cons | -0.003*** | (0.000) | $\rho^3$ | 0.017*** | (0.001) | $\gamma_1^2\rho^3$ | 0.116*** | (0.003) |
| $\rho$ | 0.058*** | (0.002) | $\gamma_1\rho$ | 0.722*** | (0.008) | $\gamma_2\rho$ | 0.732*** | (0.008) |
| $\gamma_1$ | 0.006*** | (0.001) | $\gamma_1^2\rho$ | -0.744*** | (0.007) | $\gamma_2^2\rho$ | -0.754*** | (0.007) |
| $\gamma_2$ | 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 $\hat{\theta}_{t+1}$ plotted are the results of 1,440 different combinations of $\gamma_1$, $\gamma_2$ and $\rho$ parameters in the CES production function $\hat{\theta}_{t+1} = [\gamma_1 x^\rho + \gamma_2 y^\rho + (1-\gamma_1-\gamma_2)\, z^\rho]^{1/\rho}$, where $x$, $y$ and $z$ come from 5,000 random draws from independent normal distributions.

# B Identification of the CES Function

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

# C 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.