Victimization and Skill Accumulation: The Case of School Bullying*

Miguel Sarzosa†
Purdue University

October 28, 2021

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 34% 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, Salvador Navarro, Greg Veramendi, James Heckman and Matt Wiswall. 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. The author did not have access to information leading to the identification of individuals. The data analysis was carried out in a secure server.

†Miguel Sarzosa is an assistant professor at the Economics Department in Purdue University. Email: msarzosa@purdue.edu

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. We know very little about its intermediate costs and long-term consequences. In this paper, I contribute to bridging this gap by exploring the two-way relation between bullying and cognitive and non-cognitive skills accumulation.[2] Namely, how school bullying hampers

[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 (Wang et al., 2009).

[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*—defined as relatively enduring patterns of thoughts, feelings, and behaviors that allow people to recognize and control their emotions and reactions, establish and maintain positive relationships, make responsible decisions, and set and achieve positive goals (Borghans et al., 2008; OECD, 2014)—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 *non-cognitive* and *socioemotional* skills interchangeably (Saltiel et al., 2017). Sometimes they are also referred to as *soft* skills (Heckman and Kautz, 2012).

the development of successful adults by impeding skill accumulation, and the extent 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 critical 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 that those true latent skills follow a normal distribution. Thus guaranteeing 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 essential drivers of the process. In particular, it considers the role endogenous peer victimization has on skill formation. Second, it analyzes the consequences of bullying in school in terms of skill depletion. Third, it extends my previous work on school

[3] Well-established facts about child victimization in the psychological literature inspire this two-way relation. Namely, that bullying victims suffer grave and long-lasting consequences in terms of their emotional well-being (Smith and Brain, 2000; OECD, 2017, among many others), and that the likelihood of victimization increases dramatically when the child has some behavioral vulnerability (Hodges et al., 1997; Reijntjes et al., 2010).

bullying (Sarzosa and Urzua, 2021), where I found sizable consequences borne during adulthood by providing additional insight into the channels through which 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 losses 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 an enormous burden they will carry 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 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 disruptive classmates have on their 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:

[4]The psychology and sociology literatures have been prolific in examining bullying as a social phenomenon. Among many findings, they have established 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

first, lack of adequate data; second, the non-randomness of the selection into bullying. Regarding the former, there is little longitudinal data that inquire about bullying, so researchers can use very few sources that observe individuals before and after the event. Regarding the latter, bullying's non-randomness confounds the consequences of bullying with the intrinsic characteristics that made the person a victimin the first place. The scarce existing economic literature has focused on quantifying 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 marks among Danish students. They find causal estimates by instrumenting victimization with the proportion of classroom peers whose parents have a criminal conviction. Sarzosa and Urzua (2021) 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 skills.

In this paper, I contribute to the literature of bullying by building on Sarzosa and Urzua (2021) and providing a possible explanation for the impacts they observe. While Sarzosa and Urzua (2021) estimate the Average Treatment Effect (ATE) of middle school bullying on young adult outcomes, in the present study, I elucidate one mechanism behind the creation of these gaps, showing bullying as the triggering event

(Smith et al., 1999), 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 (NAS, 2016). Bullied children generally have less self-esteem, and have a negative view of their situation (Björkqvist et al., 1982; Kochel et al., 2012). They are also more likely to feel lonely (Dake et al., 2003). These victims' characterizations highlight the importance of including explicit relation between bullying and personality throughout the analysis.

that determines a divergence in skill accumulation 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 lower than those they would have been on in the absence of victimization. In consequence, the identification strategies in the two papers differ significantly. While Sarzosa and Urzua (2021) estimates a static Roy model with unobserved heterogeneity, in the present paper I estimate a model of skill formation. Although both papers use the allocation of students to classrooms as an 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 skill development that have established that skills are dynamic and malleable (Cunha et al., 2006, 2010). 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). Skills beget skills through the natural process of getting the stock available at time t to $t+1$ and through 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] 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 meager stocks of skills (Agostinelli and Wiswall, 2016a). These dynamics give foundation to the call for early childhood development and preschool interventions (Knudsen et al., 2006; Doyle et al., 2009).

The claim that skills are malleable is backed up by a series of papers that show that skill-developing interventions modified the stock of skills of the treated population. For instance, the people treated by Perry Preschool Program have higher

[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).

non-cognitive skills—although similar levels of cognitive skills—than the 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). 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 skill levels.[6] However, evidence shows 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 in which the child grows.[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 and Bullying

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. bullying (framed as a negative investment) can hamper skill devel-

[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.

opment, and v. bullying victimization depends also on the stock of cognitive and non-cognitive skills of each person and those of his or her peers. Therefore, I propose to augment the dynamic structure in Cunha et al. (2010) to explicitly incorporate these five facts. Thus, I adopt the timeline proposed in Cunha et al. (2010) where student i that belongs to classroom $c \in C$ starts the process with an initial stock of skills of type $S \in \{A, B\}$, $\theta_{S,i\in c,\underline{\tau}}$, with A denoting non-cognitive skills and B denoting cognitive skills. Parents observe $\theta_{A,i\in c,\underline{\tau}}$ and $\theta_{B,i\in c,\underline{\tau}}$, receive an idiosyncratic investment-related shock $\varepsilon_{S,i\in c,\underline{\tau}+1}$ and decide $I_{S,i\in c,\underline{\tau}+1}$, the amount to invest between period $\underline{\tau}$ and $\underline{\tau}+1$ in each skill dimension. Analogously, $\theta_{A,i\in c,\underline{\tau}}$ and $\theta_{B,i\in c,\underline{\tau}}$ together with classroom characteristics determine the victimization that may occur between the skill measurements at $\underline{\tau}$ and $\underline{\tau}+1$, which I label $M_{i\in c,\underline{\tau}+1}$. Hence, in general, parental investment and victimization at a given moment in time $t \in [\underline{\tau}, T]$ are simultanous. Parents do not observe victimization at the time they decide their skill investment strategy. This responds to the ample evidence indicating that parents are not usually aware of their children's victimization (deLara, 2012; Waasdorp and Bradshaw, 2015; Bjereld et al., 2017; Larranaga et al., 2018).[8] Parental investment made between times t and $t+1$ respond, however, indirectly to past victimization. That is, through the effect bullying that occurred between $t-1$ and t has 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).

Skills at $\underline{\tau}+1$ are, thus, the product of the initial stock of skills $\theta_{A,i\in c,\underline{\tau}}$ and $\theta_{B,i\in c,\underline{\tau}}$, the investment decisions and the victimization that took place between $\underline{\tau}$ and $\underline{\tau}+1$

[8]Studies have shown that adolescents do not disclose their victimization to adults because they are ashamed, they underplay its consequences, or they fear their parents could make their problem worse (deLara, 2012; Larranaga et al., 2018). Other studies show that adolescents see disclosing being bullied to a parent as their last resort due to the fact that disclosure has been associated with more serious bullying experiences (Smith et al., 2001; deLara, 2008).

$I_{S,i\in c,\underline{T}+1}$ and $M_{i\in c,\underline{T}+1}$ and the realization of a skill production shock $\eta_{i\in c,\underline{T}+1}$. See a diagram with the timeline of the model's typical two-period cycle in Figure B.1 in the Appendix.

3.1 The Production Function of Skills

The empirical characterization of the skill production functions faces at least two challenges. First, all the sequences of $\{\theta_{S,i\in c,t}\}_{t=\underline{T}}^{T}$ and $\{I_{S,i\in c,t}\}_{t=\underline{T}}^{T}$ for $S \in \{A, B\}$ are not directly observable. They are latent and influence the values we observe in manifest variables such as cognitive scores, non-cognitive scales, and parental investment measures (Cunha and Heckman, 2008). Using a set of these manifest variables in place of the latent variation will severely bias the results (Attanasio et al., 2020c). Thus—as I will explain in greater detail in Section 4.1 and Appendix C—I follow (Cunha and Heckman, 2008) and Cunha et al. (2010) and arrange the numerous manifest variables available in the data in measurement systems linking the manifest variables to the latent constructs we care about. This latent factor framework takes the common variation in the measurement systems of manifest variables and disentangles the variation that comes from the unobserved factors from the one generated by random shocks and the one that comes from exogenous observable traits like gender or age. Its goal is to allow for the estimation of the distribution of the latent factors (Carneiro et al., 2003). Note that unlike recent papers like Agostinelli and Wiswall (2016a) or Attanasio et al. (2020c) that use the estimated distributions of the latent factors to approximate a value of $\left\{\widehat{\theta}_{S,i\in c,t}\right\}_{t=\underline{T}}^{T}$, and $\left\{\widehat{I}_{S,i\in c,t}\right\}_{t=\underline{T}}^{T}$ for every i, I keep the latent variables as such. A feature that will come in handy when estimating the potential outcomes model (i.e., victimization status-specific skill production functions).

The empirical characterization's second challenge is that I need to impose some functional form assumptions about the family of production functions I will estimate. The goal is to impose some parametric assumptions to make the model estimable while allowing for such flexibility that, based on the data, it allows for the recovery of a wide assortment of production functions. I consider the technology of production of skill S in period $t+1$ for those with victimization condition M_{t+1} to follow a Constant Elasticity of Substitution (CES) 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 ($I_{S,i\in c,t+1}$). 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 assumptions already outlined in Cunha et al. (2010). Second, the CES allows the inputs involved in the skill production function to have non-linear and joint effects *while allowing them to remain latent*. I am interested in one particular product of such non-linearities: the 'static complementarity' ($\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial\theta_{A,t}$ and $\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial\theta_{B,t}$). A concept introduced by Cunha and Heckman (2008) to describe how the current stock of skills affects the productivity of skill investment. I will use the same concept to analyze the *skill depleting* power of the bullying event.

The CES specification could be overly restrictive if the elasticities of substitution between inputs vary widely. In that sense, a translog function would capture non-linear effects and complementarities like the CES while allowing for different substitution parameters between inputs (Agostinelli and Wiswall, 2016a; Attanasio et al., 2020a). However, given that the translog function relies on interactions between inputs, its estimation requires inputs to be treated as observable (i.e., $\widehat{I}_{s,t+1}$ and $\widehat{\theta}_{s,t}$), which would be at odds with estimating of the model of potential outcomes I use to measure the treatment effects of victimization.[9]

[9] Recently, the CES specification has been under criticism due to its strong location and scale assumptions (Agostinelli and Wiswall, 2016b; Del Bono et al., 2020; Freyberger, 2020). I will deal with these issues in Section 4.2.2.

3.2 "Selection into Bullying"

"Selection into bullying" is non-random; it depends on the victim's and her classmates' characteristics.[10] 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 the 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. 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 ψ. Z_c is a vector containing 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).

[10]In that sense, this selection problem relates to the issues studied by 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*.

3.3 The Model of Skill Formation with School Bullying

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

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

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

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

for $S \in \{A, B\}$, where $\gamma^{M_i}_{I,S,t} = 1 - \gamma^{M_i}_{A,S,t} - \gamma^{M_i}_{B,S,t}$ for $M_i \in \{0, 1\}$, $\mathbf{1}[\cdot]$ is an indicator function that takes the value of 1 if true . $\eta^{M_i}_{S,i\in c,t+1}$ denote shocks that affect the accumulation of skill dimension S between t and $t + 1$. The CES parameters contain a superscript $M_i \in \{0, 1\}$ to indicate that the skills production functions for victimized students are different from those of non-victimized ones. 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 independent and identically distributed (iid) shocks orthogonal to contemporaneous skills, to each other, across time and across victimization condition. The mutual independence of $\eta^{M_i}_{S,i\in c,t}$, $\varepsilon_{S,i\in c,t+1}$, and $e^M_{i\in c,t+1}$ is the result of independence assumptions imposed on the measurement systems used to obtain the distributions of the latent variables $\theta_{A,i\in c,t}$, $\theta_{B,i\in c,t}$ and $I_{S,i\in c,t+1}$. I will comment further on these assumptions in Section 4.2.1. 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) (Reijntjes et al., 2010), 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 occur, allowing classmates to play different roles: victim, perpetrator, and bystanders.[11] Therefore, the question that arises is: what separates bystanders from victims? Here is where the uncommonness feature becomes essential as it operationalizes the imbalance of power bullying 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).[12]

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 identifying (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 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 every observation to have a different "treatment" to the point that, although everyone is affected by what happens inside their particular epsilon-ball, the relative position of those classmates that do not fall within it is irrelevant.

[11] 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.

[12] 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.

4 Empirical Strategy

4.1 Measurement System and Unobserved Heterogeneity

The key feature of the empirical strategy is how it deals with the fact that underlying cognitive and non-cognitive skills and investment preferences are latent rather than observable.[13] 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 measures, whose inputs include 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, ν^S).

Suppose we follow individuals for two time periods: t and $t+1$. Then, the measurement system—omitting the student and classroom subscripts to simplify notation—is the following:

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

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

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

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

where $\mathbf{T}_\tau$ is a $L \times 1$ vector that contains the scores of cognitive tests an non-cognitive measurements at time $\tau \in \{t, t+1\}$, $\mathcal{I}_{S,t+1}$ is a $L_{\mathcal{I}_S} \times 1$ vector that contains each of the

[13] 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.

investment measures made in skill $S \in \{A, B\}$ at time $t+1$. The latent variables of interest are skills $\mathbf{\Theta}_\tau = \left[\begin{array}{cc} \theta_{A,\tau} & \theta_{B,\tau} \end{array}\right]$ and investments $I_{A,t+1}$ and $I_{B,t+1}$. $\mathbf{\Theta}_\tau$ follows the bivariate distribution $F_{\theta_{A,\tau},\theta_{B,\tau}}(\cdot,\cdot)$, and $I_{A,t+1}$ and $I_{B,t+1}$ follow distributions $F_{I_{A,t+1}}(\cdot)$ and $F_{I_{B,t+1}}(\cdot)$ respectively. $\mathbf{X}_{\tau,T}$ are matrices with all observable controls affecting the scores at time $\tau \in \{t, t+1\}$, and $\mathbf{X}_{t+1,\mathcal{I}}$ is a matrix containing all observable controls affecting manifest investment measures at time $t+1$. Λ_τ^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 ν_{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 Σ_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 $\mathbf{e}_\tau^\mathbf{T} \perp \nu_{\mathbf{t+1}}^\mathbf{S}$.

Appendix C presents the arguments for the identification of the model, including coefficients, factor loadings and factor distributions. I do not impose normality to the distributions of the factors $f_{\theta_A,\theta_B}(\cdot,\cdot)$ or $f_{I_{\cdot,t+1}}(\cdot)$. 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 enables the model to replicate a wide range of distributions and 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 $\widehat{\beta}_t^T, \widehat{\Lambda}_t^T$, $\widehat{\Sigma}_{L_t}$, $\widehat{\beta}_{t+1}^{\mathcal{I}_\cdot}$, $\widehat{\alpha}_{t+1}^{\mathcal{I}_\cdot}$, $\widehat{\Sigma}_{L_{\mathcal{I}_\cdot}}$, $\widehat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$ and $\widehat{F}_{I_{\cdot,t+1}}(\cdot)$.

4.2 Estimation

4.2.1 Identification and Estimation Steps

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

$$\widehat{\xi}_{t+1} = \mathbf{T}_{t+1} - \mathbf{X}_{t+1,T}\widehat{\beta}^{T}_{t+1} = \widehat{\Lambda}^{T}_{t+1}\boldsymbol{\Theta}'_{t+1} + \mathbf{e}^{\mathbf{T}}_{\mathbf{t+1}} \tag{7}$$

Taking advantage of the orthogonality and mutually independence between $\mathbf{e}^{\mathbf{T}}_{\mathbf{t}}$, $\mathbf{e}^{\mathbf{T}}_{\mathbf{t+1}}$ and $\eta_{\mathbf{t}}$, and of the non-linearity of the skills production functions, I substitute them from (1) into the measurement system for $\widehat{\xi}_{t+1}$. For the sake of brevity, let me call $g^{M}_{S,t+1}(\theta_t, I_{t+1})$ the production function of skill S at time $t+1$ for those whose victimization condition is $M \in \{0,1\}$. Then, I can write (7) as

$$\widehat{\xi}_{t+1} = \begin{cases} \widehat{\lambda}^{\mathbf{T_A}}_{\mathbf{t+1}} g^{0}_{A,t+1}(\theta_t, I_{t+1}) + \widehat{\lambda}^{\mathbf{T_B}}_{\mathbf{t+1}} g^{0}_{B,t+1}(\theta_t, I_{t+1}) + \vartheta^{0}_{t+1} & \text{if } M_{i,t+1} = 0 \\ \widehat{\lambda}^{\mathbf{T_A}}_{\mathbf{t+1}} g^{1}_{A,t+1}(\theta_t, I_{t+1}) + \widehat{\lambda}^{\mathbf{T_B}}_{\mathbf{t+1}} g^{1}_{B,t+1}(\theta_t, I_{t+1}) + \vartheta^{1}_{t+1} & \text{if } M_{i,t+1} = 1 \end{cases} \tag{8}$$

which together with the victimization equation (9) (i.e., the empirical version of equation (2))

$$M_{t+1} = \mathbf{1}\left[\mathbf{X}_{t+1,M}\beta^{M}_{t+1} + \Lambda^{M}_{t+1}\Theta'_{i\in c,t} + \Lambda^{M_c}_{t+1}\nabla_{\psi,i\in c}(d) + \Gamma Z_{t+1,c} > e^{M}_{t+1}\right] \tag{9}$$

build a Roy-like potential outcomes model that endogenizes the bullying 'treatment', and allows me the estimation of treatment effects of victimization on skill formation (Heckman and Vytlacil, 2007). Formally:

$$ATE_{t+1}(\theta_{A,t}, \theta_{B,t}) = E\left[\theta^{S}_{t+1} | \theta_{A,t}, \theta_{B,t}, M_{t+1} = 1\right] - E\left[\theta^{S}_{t+1} | \theta_{A,t}, \theta_{B,t}, M_{t+1} = 0\right]$$

Equation (9) 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 (9), I follow Sarzosa and Urzua (2021) 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.[14]

The measurement system requires several considerations. First, note that $\vartheta^M_{t+1} = \widehat{\lambda}^{\mathbf{T_A}}_{\mathbf{t+1}}\eta^M_{A,t+1} + \widehat{\lambda}^{\mathbf{T_B}}_{\mathbf{t+1}}\eta^M_{B,t+1} + \mathbf{e}^{\mathbf{T}}_{\mathbf{t+1}}$ 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 $\left(\widehat{\lambda}^{T^l_A}_{t+1}\right)^2\sigma^2_{\eta^M_{A,t+1}} + \left(\widehat{\lambda}^{T^l_B}_{t+1}\right)^2\sigma^2_{\eta^M_{B,t+1}} + \sigma^2_{e^{T^l}_{t+1}}$ and its off-diagonal elements are of the form $\widehat{\lambda}^{T^l_A}_{t+1}\widehat{\lambda}^{T^j_A}_{t+1}\sigma^2_{\eta^M_{A,t+1}} + \widehat{\lambda}^{T^l_B}_{t+1}\widehat{\lambda}^{T^j_B}_{t+1}\sigma^2_{\eta^M_{B,t+1}}$. It is straightforward to see that Ω^M_{t+1} is identified from the fact that $\widehat{\Lambda}^T_{t+1}$ and $\widehat{\sigma}^2_{\mathbf{e}^{\mathbf{T}}_{\mathbf{t+1}}}$ are known from the first stage. Hence, I am effectively reducing the dimensionality of the computational task of estimating the model. It is now a four-dimensional unobserved heterogeneity problem: two dimensions of skills at t and the investment latent factor for each skill. Second, identification of this potential outcomes model and its associated treatment parameters requires that $e^M_{t+1} \perp \left(\vartheta^0_{t+1}, \vartheta^1_{t+1}\right)$ (a modified version of

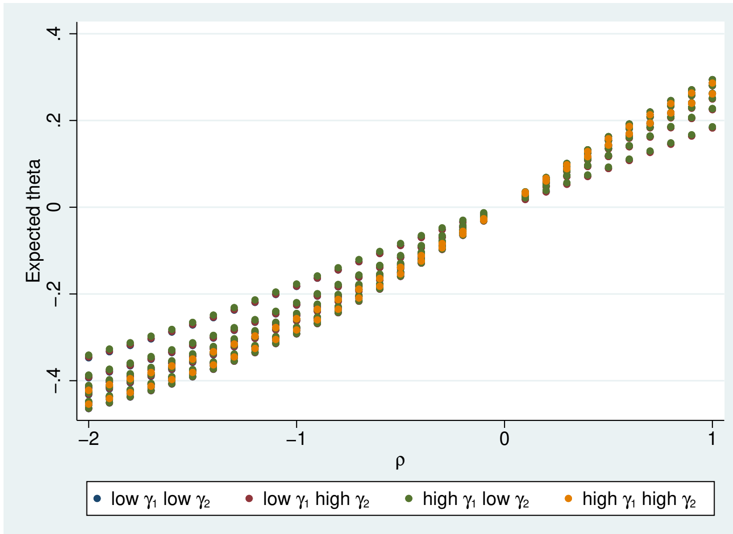
[14]See the details of these two variables in Sarzosa and Urzua (2021). The *family violence* measure comes from the following questions: 1. I always get along well with brothers or sisters, 2. I frequently see parents verbally abuse each other, 3. I frequently see one of my parents beat the other one, 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 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).

assumption (A-1c) in Heckman et al., 2016). This assumption implies two underlying assumptions: $e^M_{t+1} \perp \mathbf{e^T_{t+1}}$ and $e^M_{t+1} \perp \left(\eta^0_{A,t+1}, \eta^0_{B,t+1}, \eta^1_{A,t+1}, \eta^1_{B,t+1}\right)$. The former is a mild assumption as its violation would require a very unique kind of shock. One that jointly shifts the chances of victimization and scores recorded by the cognitive tests and non-cognitive measures in $t+1$, but *does not affect* the stock of skills at that moment in time. The latter assumption—which I alluded to in Section 3.3—maintains independence of the shocks to the chances of victimization and the part of the variation in the latent variable $\theta_{S,t+1}$ not explained by $\theta_{A,t}$, $\theta_{B,t}$ and $I_{S,t+1}$. Therefore, shocks that simultaneously alter the 'sorting into victimization' and period $t+1$ skills without going through $g^M_{S,t+1}(\theta_t, I_{t+1})$ are considered threats to the identification of the model. To deal with this concern, I estimated a version of the model that includes numerous shocks that could have the potential of violating the identifying assumption (i.e., death of parent, parent failed in business, parent lost job, parent was hospitalized) both in the victimization equation (9) and in the potential outcome equations (8) as an additional observable control. Table 8 in the Web Appendix shows that although the shocks are significant determinants of victimization, the parameters of the production function remain unaltered. The fact that the results are robust to the introduction of the shocks illustrate that their scope is too small to be meaningful.

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 $S \in \{A, B\}$. These normalizations are at odds with the fact that $E\left[\widehat{\theta}_{\cdot,t+1}\right]$ shifts with changes in ρ, as shown in Figure 1. It simulates 1,440 different combinations of γ_1, γ_2 and ρ to generate $\widehat{\theta}_{t+1} = \left[\gamma_1 x^\rho + \gamma_2 y^\rho + (1-\gamma_1-\gamma_2) z^\rho\right]^{1/\rho}$, where x, y and z come from 5,000 random draws from independent normal distributions. Figure 1 evidences that estimating a model that fits $E\left[\widehat{\theta}_{\cdot,t+1}\right] = 0$ greatly constrains the set of possible values that $\widehat{\rho}$ can take, and its combinations with the other parameters in the CES function. In other words,

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



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

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 $E\left[\widehat{\theta}_{\cdot,t+1}\right]$ and the CES parameters is predictable as evidenced by Figure 1. In fact, Table D.1 in the Appendix shows that a flexible cubic polynomial in the CES parameters (i.e., $\mathbf{P}_3\left(\gamma_1, \gamma_2, \rho\right)$) captures 99.98% of the variation of $E\left[\widehat{\theta}_{\cdot,t+1}\right]$. Hence, in order to avoid $E\left[\widehat{\theta}_{\cdot,t+1}\right] = 0$ constraining the possible values of the CES parameters, I use $\mathbf{P}_3\left(\gamma_A, \gamma_B, \rho_\cdot\right)$ as a shifter of the mean of $\widehat{\theta}_{\cdot,t+1}$ during estimation. That way, it counters the mean-shifting that mechanically occurs when $\widehat{\rho} \neq 0$. In practice, I am allowing $E\left[\vartheta^M_{t+1}\right] = -\widehat{\alpha}^{T_A}_{t+1}\widehat{\mathbf{P}}_3\left(\gamma^M_{A,A}, \gamma^M_{A,B}, \rho^M_A\right) - \widehat{\alpha}^{\mathbf{T_B}}_{\mathbf{t+1}}\widehat{\mathbf{P}}_3\left(\gamma^M_{A,B}, \gamma^M_{B,B}, \rho^M_B\right)$.

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.[15]

4.2.3 The problem of joint causality

The empirical model presented so far relies on the assumption that scores at t are measured before any victimization has occurred. However, given the survey's timing (it takes place by the end of the school/calendar year), cognitive scores and non-cognitive measures were collected after some victimization had already happened. This may cause a problem of joint causality analogous to the one addressed by Hansen et al. (2004) when exploring the relationship between skills, manifest scores, and schooling at the time of measurement. They face the simultaneity issue because schooling develops skills and boosts test scores, and also high skilled people find it easier to achieve higher schooling attainment. Hansen et al. (2004) show that, by recognizing that the same unobserved skills determine both schooling and scores, they can overcome the joint causality problem and identify the distributions of those skills.

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 the manifest measures observed in the first survey wave are generated by the initial unobserved skills. Therefore, using Hansen et al. (2004) framework, I can disentangle skills, manifest measures and victimization. To do that, I will extend the structure of the measurement system in (3) to incorporate the one proposed by Hansen et al. (2004).[16]

[15] Urzua (2008) shows that—under mild linearity assumptions in measurement systems (3) 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 identified without normalizing $E[\theta_{S,t+1}] = 0$.

[16] 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,

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

$$\mathbf{T}(M_t) = \mathbf{X}_{t,T}\beta^{T}(M_t) + \Lambda^{T}(M_t)\,\mathbf{\Theta}'_t + \mathbf{e^T}(M_t) \tag{10}$$
$$M_t = \mathbf{1}\left(\mathbf{X}_{t,M}\beta^{M}_t + \Lambda^{M}_{\tau t}\Theta'_{i\in c,t} + \Lambda^{M_c}_t \nabla_{\psi_t, i\in c}(d) + \Gamma Z_{t,c} > e^{M}_t\right)$$

Note that this implies that the matrices β^T_t and Λ^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 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" regulating 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,

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.

Table 1: Descriptive Statistics

Total sample size	3,449			*Incidence of Bullying*	
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. The income figure shown corresponds to the median monthly per-capita household income.

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 has a sampling scheme that is critical for identifying 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 identifying 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.[17] Each wave, information was collected in two separate questionnaires: one for the teenager and another for the parents or guardians.

Another critical feature of the KYP-JHSP regarding this study is that it collects very detailed information on personality traits and behavioral responses through a comprehensive battery of personality questions consistent across waves. The KYP-JHSP inquires about academic performance, student effort, and participation in dif-

[17] 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.

ferent kinds of private tutoring. The survey also asks about time allocation, leisure activities, social relations, attachment to friends and family, participation in deviant activities, and victimization in different settings, including bullying. While the survey often asks the children about their parents' involvement in many aspects of their lives, parents and guardians answer only a short questionnaire covering household composition and their education, occupation and income.

As with all other personal characteristics collected in the KYP-JHSP, bullying 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 during the last year. Given that the KYP-JHSP collects its data during late November and the Korean school year runs from March to December, one can interpret the question as asking for bullying events during the school year that is about to end.

Even though psychologists define bullying to include more than physical violence (see its definition in the Introduction of this paper), due to the wording of the question in the KYP-JHSP, the kids in the study respond to its most direct and less subtle versions of bullying.[18] This way of reporting about bullying is in line with the findings in several international studies that find that children "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).[19] 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 the incidence—and its year-to-year decline—reported in international studies (OECD, 2017; Scheithauer et al., 2006; Ryoo et al., 2015). 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). For this study, I use the bullying measured in waves one, two, and three.[20]

[18]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.

[19]See, for instance, Madsen (1996); Smith et al. (2002).

[20]The KYPS-JHS collects information about the incidence of bullying (i.e., a dichotomous variable)

Data and institutional requirements aside, it is worth noting that—like in the US and many other countries globally—bullying is a critical issue in the South Korean society, usually characterized by ultra-competitive academic environments that praise scholastic achievement.[21] Not surprisingly, such environments foster unhappiness and aggressiveness in the classrooms, a fertile ground for bullying. Given the link between bullying and suicides (Kim and Leventhal, 2008; Kim et al., 2009), and the striking suicide rate among young people in South Korea,[22] the government has deployed active policies aimed at curving these phenomena.[23]

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

As explained while describing the empirical strategy in Appendix C, estimating the latent heterogeneity's distribution parameters requires at least three manifest measures per factor. In this subsection, I present how I constructed those measures for each dimension of the unobserved heterogeneity.

5.1.1 Cognitive Scores

The KYP-JHSP contains information on grades and academic performance. In particular, I use two self-reported measures on the students' achievement in: i.) math and science, and ii.) language (Korean) and social studies, together with the the score

---

and about its frequency. However, the reported frequency has very little variation. This may stem from the fact that bullying—by definition—implies a repetitive behavior. So, children might report multiple attacks under one bully-bullied relation.

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

[22] 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/).

[23] 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.

obtained in a comprehensive test taken at the end of the academic year. The exam is considered high stakes—the scores matter for future applications to high school. Students aiming to enter high-achieving high schools, which will later springboard them to top universities, need to get top marks in these exams consistently.[24]

Appendix C indicates that one requirement for identifying the parameters associated with the correlated latent factors and the adjunct measurement system is to have at least one exclusive measure per factor dimension. That means that there must be at least one cognitive measure whose production function does not include non-cognitive skills. Of the three cognitive scores, two of them are course achievement measures, and one is an exam score. Previous literature has shown that course grades may not be orthogonal to non-cognitive skills (Heckman et al., 2011). Course grades—being the summation of multiple tasks throughout the school year, including homework and assessments often relating to classroom behavior—are, to a significant degree, the product of non-cognitive skills. Thus, the production function of course grades must be modeled using both cognitive and non-cognitive skills as inputs. As shown in Section 4, my model considers this feature of the data and incorporates it into the estimation by allowing the math and science, and the language and social studies grades the be affected by both skill dimensions.

The yearly exam, on the contrary, is a one-shot assessment and, thus, less dependent on non-cognitive skills than course grades. Indeed, children who did their homework and behaved well throughout the year are more likely to have learned more. However, the yearly test does not measure those behaviors directly as course grades do. In fact, wave one correlations between the three cognitive scores and a factor collecting the common variation in non-cognitive measures via a principal component analysis show that the yearly exam score is orthogonal to non-cognitive variation (0.016, not statistically different from zero). In contrast, the correlations

[24]Compulsory schooling in South Korea finishes at the end of middle school. However, we should note that 99.7% of middle school graduates continue their education into high school. In 2010, the high school graduation rate in South Korea reached 94%, the highest among OECD countries (OECD, 2012).

between grades and the non-cognitive variation are statistically significant (0.122 for math and 0.103 for language, both statistically different from zero). Based on this evidence, I choose the yearly test as the exclusive measure for the identification of cognitive skills.

5.1.2 Non-Cognitive Measures

To identify non-cognitive skills, I use measures of locus of control, responsibility, and self-esteem. The KYP-JHSP records the socio-emotional information in categories that group the respondent's reactions 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.[25] This method makes the manifest scores more continuous, which is essential for the estimation procedure.

Regarding the choice of the dedicated non-cognitive measure required for the identification of the correlated skills, I choose the measure that correlates the least with a factor collecting the common variation in the cognitive scores via a principal component analysis. The correlation between self-esteem and the cognitive variation is less than *a fifth* of the correlations between the cognitive variation and the other non-cognitive measures. A one standard deviation increase in cognitive skills is associated with an increase in the self-esteem score of only 4.4% if a standard deviation. These results provide evidence in favor of using self-esteem and not any of the other non-cognitive measures as the dedicated manifest variable for the identification of the

[25]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.

latent non-cognitive skills factor.

Notably, the yearly test score and self-esteem-–the two manifest variables chosen to be the dedicated measures for each skill dimension-–are the manifest variables with the lowest piece-wise correlation among all the possible pairs.

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 parental harmony. The first measure indicates how often the parents beat, physically hurt, yell at or inappropriately addressed the child. Parental control relates to how well parents know where the kid is, who she is with, what she is doing, and when she is returning home. Parental harmony collects information related to the level of care and interest in her life the kid feels from her parents.[26]

The measures used to identify the cognitive skill investment factor relate to each kid's enrollment in private tutoring. South Korean society gives enormous importance to academic success. South Korean's out of pocket expenditures on education amount 0.8% of the GDP—more than two times the OECD average (Choi and Choi, 2015). Hence, it is not uncommon for kids to enroll in after-school academic programs. By age 14, around four-fifths of the sample attend some tutoring. Thus, as manifest variables of cognitive skill investment, I use a scale of how personalized the tutoring sessions are,[27] the time spent in tutoring, and the tutoring cost.

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

[27]This manifest score collects information on 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.

6 Results[28]

As explained in Section 4.2, estimation was divided into two stages.[29] For that reason, I used the Limited Information Maximum Likelihood (LIML) technique to correct the second stage's standard errors (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 per-capita household monthly income. Below, I will explain the additional variables specific to each equation (i.e., exclusion restrictions).

6.1 Model Fit

In the first step, I estimate the initial distribution of skills (age 14) from model (10), which incorporates the structure proposed by Hansen et al. (2004) to address the possible problems of joint causality. Figure 1 in the Web Appendix presents the estimated initial distribution of skills. As expected non-cognitive and cognitive skills are positively correlated: $corr\left(\theta_{A,t}, \theta_{B,t}\right) = 0.450$.

Table 2 and Figures 2 show that the model fits extremely well the actual data. The former shows that the model matches the incidence of bullying almost exactly, and that the means and standard deviations of the simulated scores are very close to the ones obtained from the actual cognitive and non-cognitive measures for each victimization state (i.e., bullied or not bullied). I cannot reject the null of equality of means in any of the 12 cases.

Figure 2 plots the predicted values of the manifest variables provided by the

[28]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.

[29]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. Thus, kids with high levels of one skill tend to have high levels of the other skills. 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. A 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.

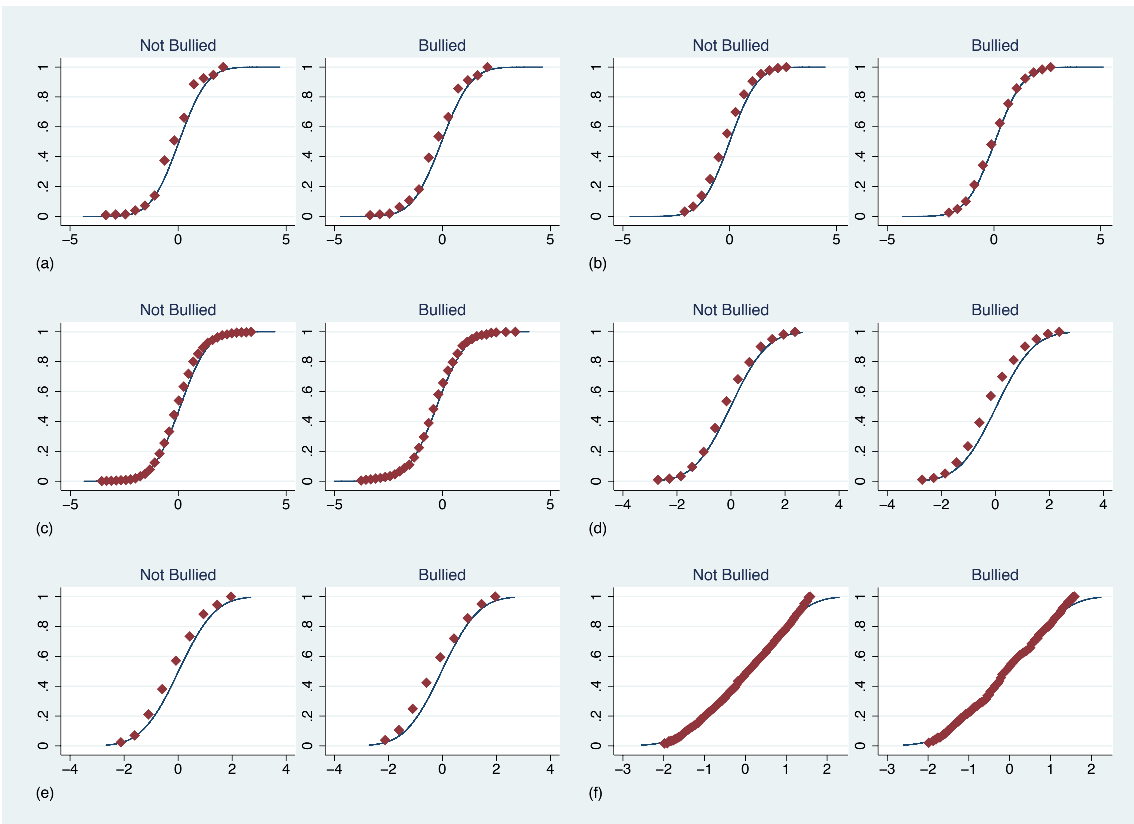
Table 2: Goodness-of-fit of the model

	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`. stands 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` stands for the year exam score. The predicted values come from simulations based on the estimated parameters of the model. *K-S p-value* reports the probability of rejecting Kolmogorov–Smirnov test's null hypothesis of equality of distributions between the predicted and the actual (reference) samples.

Figure 2: Actual vs. predicted scores cumulative distributions conditional on victimization at $t = 1$



Note: Actual (diamond) and predicted (line) cumulative distributions plotted of the following manifest variables: (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

model against the actual CDF of each cognitive test or non-cognitive measurement observed in the data. The figures show a remarkable fit in all scores, regardless of the victimization condition. I corroborate this by performing a Kolmogorov-Smirnov test on the predicted and actual measurement distributions. The results 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 measurement for which I fail the K-S test is self-esteem, even though the model closely matches its first and second moments. 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. When I smooth the observed distributions of self-esteem using kernel approximations, the K-S test statistics get closer to the non-rejection values.

6.2 Results from the Model of Skill Formation

6.2.1 Incidence of victimization

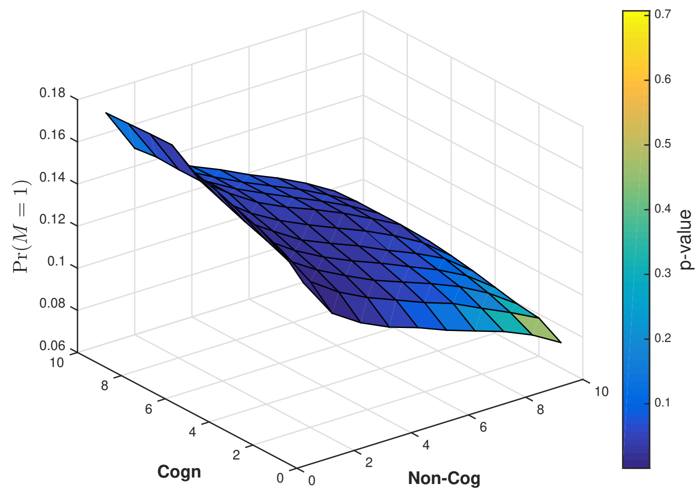
Column 1 in Table 3 shows the relation between skills and selection into bullying. Kids with less non-cognitive skills are significantly more likely to be bullied. A one standard deviation decrease in non-cognitive skills increases the likelihood of being victimized by 2.26 percentage points ($-0.071 * \sigma_{\theta_{A,t}}$, where $\sigma_{\theta_{A,t}} = 0.308$). It represents an increase in the probability of being victimized by about a fifth. Column 1 in Table 3 also shows the importance of the relation between own and peer characteristics has in determining peer victimization. Controlling for their observable characteristics and skill levels, kids placed in a school in where their non-cognitive skills are uncommon are significantly more likely to be bullied. The results indicate that the average student's victimization likelihood drops by one percentage point with each additional classmate with similar non-cognitive skill endowments. Interestingly, uncommonness in terms of income also encourages victimization. Bullying probability falls by half a percentage point for each additional classmate with a family income level similar

Table 3: Estimating the Model of Skill Formation

	(1)	(2)	(3)		(4)	(5)	(6)	(7)
					$M_{t+1}=0$		$M_{t+1}=1$	
	M_{t+1}	$I_{A,t+1}$	$I_{B,t+1}$		$\theta_{A,t+1}$	$\theta_{B,t+1}$	$\theta_{A,t+1}$	$\theta_{B,t+1}$
$\theta_{A,t}$	-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)
$\theta_{B,t}$	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)
$\nabla\left(\widehat{\theta}_{A,t}\right)$	-0.007***			ρ	-0.084	-0.032	0.357	-0.196
	(0.003)				(0.084)	(0.042)	(0.347)	(0.145)
$\nabla\left(\widehat{\theta}_{B,t}\right)$	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 (9) 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 14). $\nabla(\cdot)$ refers to the number of classmates within a window of 10% of a SD around observation i. $\widehat{\theta}_A$ is the residualized measure of self-esteem and $\widehat{\theta}_B$ 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 $\theta_{B,t}$ (i.e., $\gamma_{B,t}$) can be obtained from $\gamma_{B,t} = 1 - \gamma_{A,t} - \gamma_{I,t}$.

Figure 3: Probability of Being Bullied



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

to the one of the prospective victim. These results are in line with the psychological literature that links victimization with those considered weird or unlikeable (e.g., Hodges et al., 1997), and remarkably robust to the inclusion of the percentage of classmates that come from troubled families and the percentage of bullies in the classroom.[30]

The fact that the model relies on identifying unobserved heterogeneity allows me to quantify the victimization probability for the average student and for every combination of skills at a given point in time. Figure 3 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 than the average student. In addition, Figure 3 shows that among those with low non-cognitive skills, the ones with higher cognitive skills are three percentage points more likely to be victimized than those at the bottom of the cognitive skill distribution. These results reflect the

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

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 process of skill formation between ages 14 and 15.[31] Two main results stand out: the massive importance of self-productivity and the relatively low productivity of parental investments in skill development.

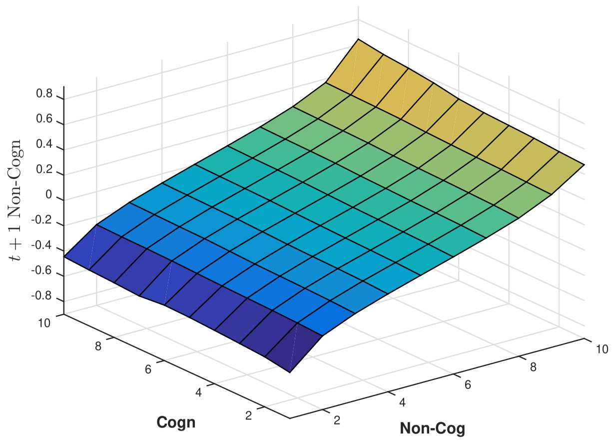
Self-productivity. Figures 4a and 4b show that high non-cognitive skills produce high future non-cognitive skills and that marginal increments of those initial skills are very productive (i.e., non-cognitive skills self-productivity $\partial\theta_{A,t+1}/\partial\theta_{A,t} > 0$ for the entire $(\theta_{A,t}, \theta_{B,t})$ space). Table 3 shows that the non-cognitive skills' input shares in the production of future cognitive skills amount to 0.904 and 0.952, depending on the victimization status. These results align well with the estimates found in existing literature.[32] Cunha et al. (2010) report that input share to be 0.868 among

[31]Point estimates of ρ suggest that the production of skills among *non-victims* follows a Cobb-Douglas specification (i.e., $\rho \approx 0$). That does not seem to be the case for *victims*, especially in their non-cognitive skills, where the point estimate reach $\rho = 0.357$. Although they are not statistically different from zero, the point estimates highlight one advantage of the method introduced in Section 4.2.2: estimating a Cobb-Douglas is only one of the possible results of the estimation. The fact that they are not statistically different from zero could be due to a lack of power as victims comprise only 11% of the sample, and the model—relying on unobserved heterogeneity and non-linear functions—is data-intensive.

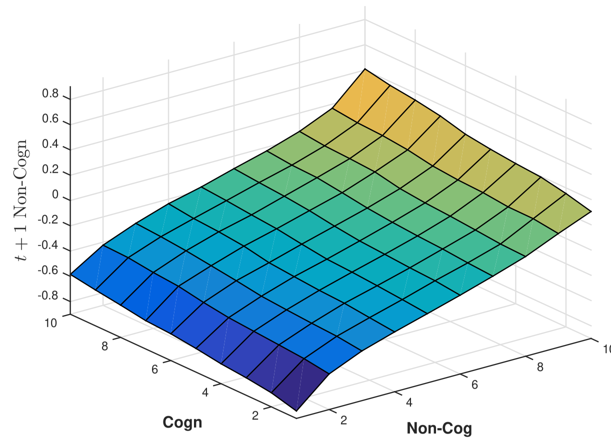
[32]Comparing estimates of skill production functions with those available in the literature can be difficult due to the vast differences in contexts and ages of the subjects on which researchers have data. The closest contexts to the one I analyze are those in Cunha et al. (2010) and Agostinelli and Wiswall (2016a). They analyze data from a developed country: the US. Also, the students' age ranges in their studies are close to the age range in my sample—although they do not overlap. Cunha et al. (2010) follows children ages 7 to 13 and Agostinelli and Wiswall (2016a) estimates a model of skill formation of children ages 11 years-old. Other existing papers study very young children in developing countries (see Attanasio et al., 2017, 2020b,c). Notwithstanding the significant differences, even in those contexts, papers still find evidence of high self-productivity of skills and the low productivity of parental investments.

Figure 4: $\theta_{S,t+1}$ as a function of $\theta_{A,t}$ and $\theta_{B,t}$

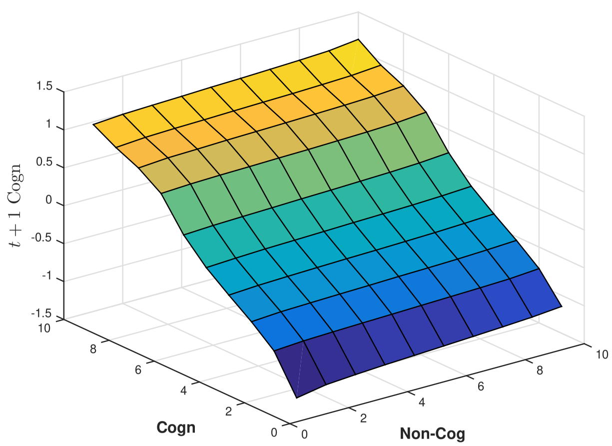
(a) Non-Cognitive: $M_{t+1} = 0$



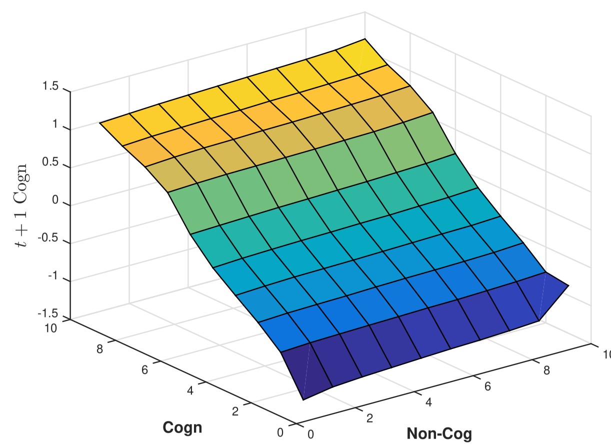
(b) Non-Cognitive: $M_{t+1} = 1$



(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 model of skill formation.

white American children between the ages of 7 and 13. Likewise, Figures 4c and 4d show that cognitive skills production relies heavily on past levels of cognitive skills. My estimates indicate that the cognitive skills' input shares in the production of future cognitive skills among Korean adolescents is 0.838 and 0.864, depending on the victimization status. Cunha et al. (2010) and Agostinelli and Wiswall (2016a) report that input share to be 0.902 and 0.910 among America pre-adolescents.

Cross-productivity. Figures 4a and 4b also demonstrate that cognitive skills are unimportant in the non-cognitive skill production process except that higher initial cognitive skills make the marginal increments of the initial non-cognitive skills more productive (i.e., $\partial^2\theta_{A,t+1}/\partial\theta_{A,t}\partial\theta_{B,t} > 0$). Likewise, Figures 4c and 4d show that although the existing levels of non-cognitive skills contribute to the cognitive skills' production process, their contribution is small compared to that of the existing cognitive skills stock. For instance, going from decile one to decile ten in the non-cognitive skills distribution has the same effect on the production of cognitive skills as increasing the cognitive skills input by one decile.

Productivity of investment. Table 3 indicates parental investments are relatively unimpactful in the production of skills.[33] The parental investment's input shares range between 0.03 and 0.057. These meager input shares among older children are also found in Cunha et al. (2010) (0.02 and 0.055 in the cognitive and non-cognitive skills' production functions) and Agostinelli and Wiswall (2016a) (0.087 in the production of the cognitive skills).

My results indicate 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 investment choices do not reverse this path dependence. Columns 2 and 3 in Table 3

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

Table 4: ATE of Being Bullied on Next Period Skills

	$\theta_{A,t+1}$		$\theta_{B,t+1}$	
	$\mathbb{E}_{\theta_t}\left[ATE(\theta_t)\right]$	$ATE(\theta_t = \bar{\theta}_t)$	$\mathbb{E}_{\theta_t}\left[ATE(\theta_t)\right]$	$ATE(\theta_t = \bar{\theta}_t)$
Estimated	-0.249***	-0.257***	-0.009	-0.006
	(0.020)	(0.020)	(0.019)	(0.019)
As SD of $\theta_{S,t+1}$	-0.399	-0.413	-0.007	-0.005

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

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 pass their high stock on to the next period and are more prone to invest in their development.[34]

6.2.3 Effects of Bullying on Skill Production and Future Bullying

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, given a particular level of period t skills. That is, $ATE\left(\theta_{A,t},\theta_{B,t}\right) = E\left[\widehat{\theta}_{S,t+1}|\theta_{A,t},\theta_{B,t},M_{t+1}=1\right] - E\left[\widehat{\theta}_{S,t+1}|\theta_{A,t},\theta_{B,t},M_{t+1}=0\right]$ for $S \in \{A,B\}$. In Table 4, I present two summarizing estimates of the effect bullying has on skill accumulation. First, I present the mean average treatment effect: $\mathbb{E}_{\theta_t}\left[ATE(\theta_t)\right] = \int ATE(\theta_t)dF(\theta_t)$, where I aggregate the treatment affects across all

[34] In Section 3.3 of the Web Appendix, I present the results of a model where non-cognitive investments directly affect the production function of cognitive skills. The results do not differ from the ones presented in Table 3. If anything, the share parameters of non-cognitive investment on cognitive skill development are even smaller than the investment share parameters estimated in the main model where a distinct cognitive investment factor affects cognitive skill production.

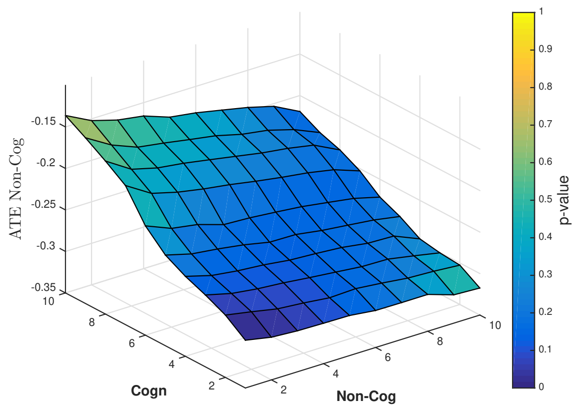
levels of period t skills. Second, I show the average treatment effect for the average student: $ATE\left(\bar{\theta}_{A,t}, \bar{\theta}_{B,t}\right) = E\left[\widehat{\theta}_{S,t+1}|\bar{\theta}_{A,t}, \bar{\theta}_{B,t}, M_{t+1} = 1\right] - E\left[\widehat{\theta}_{tS,+1}|\bar{\theta}_{A,t}, \bar{\theta}_{B,t}, M_{t+1} = 0\right]$ for $S \in \{A, B\}$, where $\bar{\theta}_{S,t}$ represents skill S mean. I find that, on average, 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, 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 losses 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 increases of 48.77% of a standard deviation in the depression symptom scale, 38.1% of a standard deviation in the levels of stress caused by insecurities regarding his or her image, and a third of a standard deviation in the levels of stress caused by issues regarding school.[35]

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. Note that the fact that bullying does not affect the accumulation of cognitive skills could be due to the nature of the victimization itself or that cognitive skills are *less malleable* than non-cognitive skills during adolescence (Walsh, 2004; Kautz et al., 2014). Then, as a robustness check, I estimate a version of the model in which cognitive skills are allowed to evolve but are not subject to the effects of victimization. The results collected in Web Appendix 3.2

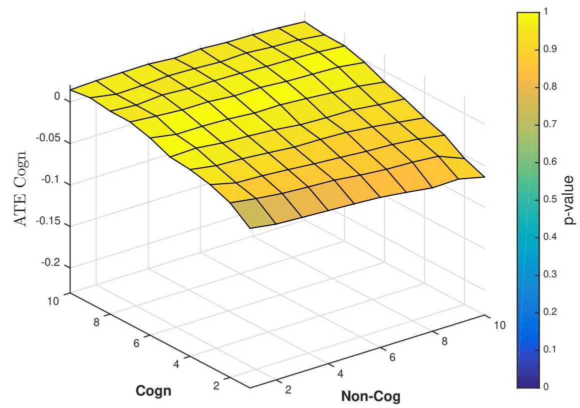
[35] In Section 2.4 of the Web Appendix, I present detailed result of estimating models of unobserved heterogeneity at age 16 of the form $Y = \mathbf{X}_Y \beta^Y + \alpha^{Y,A}\theta_A + \alpha^{Y,B}\theta_B + e^Y$, where Y is depression, stress in different situations, and the likelihood of smoking,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

Figure 5: $E\left[\theta_{S,t+1}|\theta_{A,t},\theta_{B,t},M_{t+1}=1\right]-E\left[\theta_{S,t+1}|\theta_{A,t},\theta_{B,t},M_{t+1}=0\right]$

(a) Non-Cognitive



(b) Cognitive



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

show that the impact of victimization on non-cognitive skill accumulation remains unchanged.

Analyzing the effect beyond the mean. Figure 5a presents the effect of bullying on the next period non-cognitive skills for each initial skills level. It shows that the kids who suffer the greatest negative impact come into the process with low stocks of skills. Victims with low levels of skills lose almost half of a standard deviation of non-cognitive skills, while victims with high stocks of skills lose 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 be those with 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.[36] Table 5 attests to

[36]In Web Appendix 4, I explore a different source of heterogeneity in the consequences of being bullied. I estimate a model that allows for different production functions depending on the number

Table 5: Decile of $\theta_{A,t+1}$ and victimization probability in $t+2$ that students would end up facing if victimized in $t+1$, by skills decile in t

$Q_{10}(\theta_{A,t})$	1	2	3	4	5	6	7	8	9	10
$\Pr(M_{t+1}=1)$										
	0.153	0.136	0.128	0.122	0.117	0.112	0.107	0.102	0.096	0.085
$E[\theta_{A,t+1} \vert M_{t+1}=1]$										
	-0.751	-0.542	-0.433	-0.346	-0.269	-0.189	-0.108	-0.014	0.109	0.344
$Q_{10}(\theta_{A,t+1})$	1	1	1	1	2	3	4	5	7	9
$\Pr[M_{t+2} \vert M_{t+1}=1]$										
	0.087	0.078	0.731	0.070	0.067	0.064	0.062	0.056	0.055	0.048

Note: $Q_{10}(x)$ stands for decile of x. Estimations obtained from 40,000 simulations based on the parameter estimates of the model of skill formation. Unconditional probability of victimization in $t+1$ is 11.2% and in $t+1$ is 4.62.

that. It shows how bullying shifts students to lower deciles of the next-period skills distribution. 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 who start with abundant stocks of skills fall closer to their original place in the skills distribution. Victims from the top decile at t end up in the ninth decile at $t+1$.

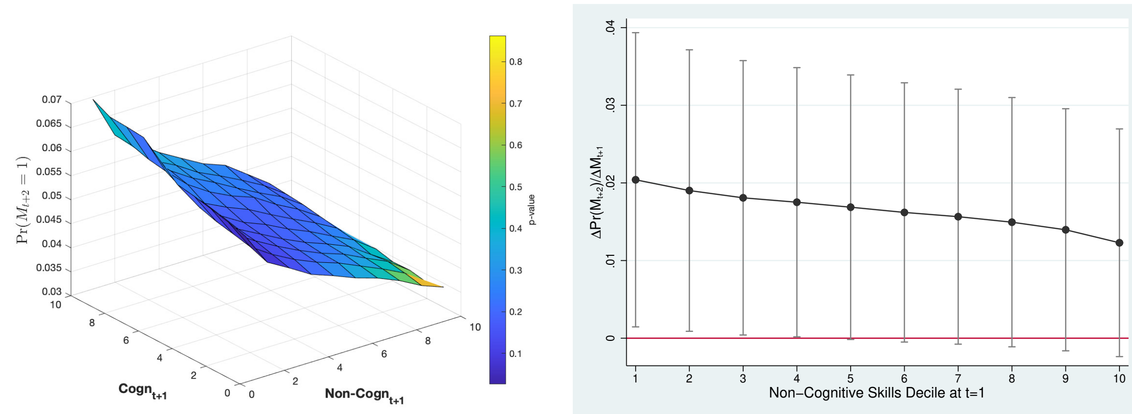
Such skill depletion between t and $t+1$ due to bullying increases the chances of being bullied again in $t+2$. The bottom row of Table 5 shows evidence of that. The likelihood of being bullied in $t+2$ for those bullied in $t+1$ exceeds 4.62—the

of bullies in the classroom. I find that the negative impact on non-cognitive skill development of being bullied is larger in classrooms with lower fractions of perpetrators. I also find that the ATE on the students in classrooms with a lower fraction of bullies has a steeper gradient with respect to the initial level of non-cognitive skills than in classrooms with a high fraction of bullies. A logic that considers that the sense of desperation might differ depending on the context where the victimization is taking place can explain these results. Classrooms with a higher fraction of bullies have more victims. Thus, a victim in a high-bullying classroom has many peers going through the same as her, while a victim in a low-bullying classroom could feel a greater sense of desperation as she will feel she is more of a target.

Figure 6: Victimization in $t+2$

(a) Probability of Being Bullied in $t+2$ by skill levels at $t+1$

(b) ATE of Being Bullied in $t+1$ on the Likelihood of Being Bullied in $t+2$ by Skill Levels in t



Note: Results based on 40,000 simulations based on the estimated parameters of the model of skill formation. Panel (b) presents how the probability of being victimized at $t+2$ changes due to having being victimized in $t+1$ for every initial level of non-cognitive skills. Namely, $E\left[M_{t+2}|M_{t+1}=1,\theta_{A,t}\right]-E\left[M_{t+2}|M_{t+1}=0,\theta_{A,t}\right]$. The spikes represent the 90% confidence intervals.

unconditional probability of being victimized at that period. Using the dynamic features of my model, I can calculate the ATE of prior bullying on the chances of being victimized again. To do so, the model exploits two facts: i) that victimization in $t+2$ depends on $t+1$ skills, as indicated in Figure 6a; and ii) that those victimized have their skills $t+1$ depleted. In consequence, I find that those bullied in $t+1$ are, on average, 1.65 percentage points more likely to be bullied again next period. That effect is not only statistically significant but economically meaningful. It represents a massive 34.6% increase relative to the overall victimization incidence in $t+2$. When disaggregating the effect by the initial level of skills, Figure 6b shows that the effect is significant for the students who start the process with relatively low skills. For instance, for students whose initial skill endowments place them in the first decile of the non-cognitive skill distribution, being bullied in $t+1$ increases the probability of being victimized in $t+2$ by 2.04 percentage points. That amounts to a 42.8% increase relative to the overall victimization incidence in $t+2$.

The channel through which these effects materialize is, of course, skill depletion. Low non-cognitive skilled students are more likely to be bullied in $t+1$. Due to that, they accumulate fewer non-cognitive skills during that period relative to what they would have if they had not been bullied. Now, with substantially less non-cognitive skills, they face a higher probability of being victimized in $t+2$. These results show the importance of the model's dynamics. Even though the overall incidence of bullying drops dramatically from year to year, victimization becomes more selective (as described in the psychological literature by Nylund et al. (2007) and Reijntjes et al. (2010)). Those who end up being bullied are most likely those who were bullied before.

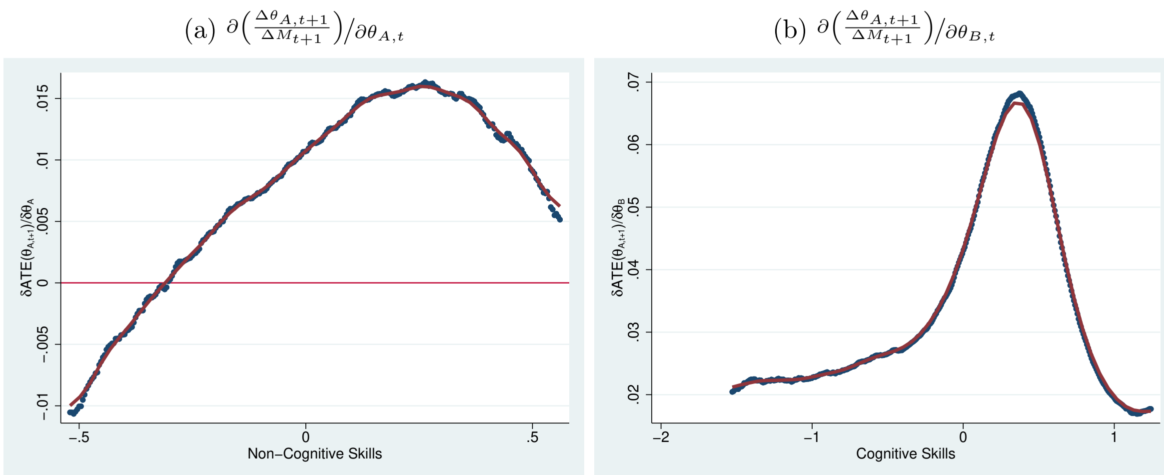
All the evidence presented in this paper confirms the existence of a self-reinforcing mechanism: kids who start the process with low levels of skills are more likely to be bullied and thus have their stock of skills depleted. These forces send them in a downward spiral by making them even more at risk of being victims of bullying in the future.[37] Subsequent bullying events will be much more harmful, preventing them from acquiring the non-cognitive skills they lack.

6.2.4 Complementarities

As explained in Section 3, an essential feature of the model is that it allows the analysis of complementarities between skills and bullying. Namely, the measurement of how much a marginal change in previous period skills modifies bullying's effect (i.e., $\partial\left(\frac{\Delta\theta_{S,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{S',t}$ for $S=\{A,B\}$ and $S'=\{A,B\}$). According to Figure 7a, marginally increasing the initial levels of non-cognitive skills will result in small reductions in bullying's negative effect on future period skills. This result attests to the fact that the impact of bullying is relatively constant across the entire non-cognitive skills distribution. 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.

[37] The notion of a vicious cycle between emotional and behavioral problems and victimization has been explored in psychology. See Reijntjes et al. (2010) and Bowes et al. (2013).

Figure 7: Static Complementarity



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

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, those adverse effects would shrink 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.

Even though I showed that investment in skills during middle school years is often unproductive, the static complementarity results suggest that even a tiny bit of skill accumulation during earlier years 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

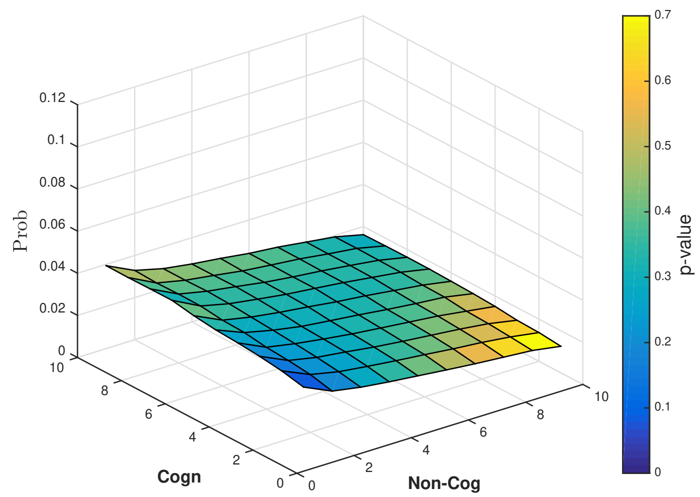
Several anti-bullying campaigns have been deployed all around the world in an ambitious effort to eliminate this unwanted phenomenon.[38] My findings indicate there are at least two fronts on which policymakers can work. First, the development of non-cognitive skills. Non-cognitive skilled kids will be less likely to be victimized. Moreover, if they happen to be bullied, the impact on their skill accumulation path is much lessened. The strong dependence of current skill levels on past skill levels heightens the importance of developing non-cognitive skills at young ages.

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 finding leads to a 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 with similar stocks of non-cognitive skills, as measured by the self-esteem score. This exercise ignores geographical distances. It sorts the universe of students with respect to their self-esteem scores and split them into classrooms according to the typical classroom size in South Korea.

Figure 8 presents the results of these simulations. As in Figure 3, it plots the likelihood of being bullied for every skill level. A comparison between these two figures shows the massive impact of reducing in-classroom non-cognitive skill heterogeneity 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 a probability

[38] 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 model of skill formation.

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

Of course, this exercise ignores all other possible consequences that the homogenization of classrooms along skill lines might have. Being in a skill-diverse classroom might be beneficial to students—in particular, those not victimized—in other domains. My simulation cannot specify whether the benefit of reduced bullying due to the homogenization of classrooms outweighs the potential positive implications of a skill-diverse classroom. The exercise shows that *one* implication of homogenization *along skill lines* is less bullying.

8 Conclusions

This paper develops and estimates a structural model of skill accumulation that introduces endogenous social interactions as drivers of the skill formation 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. In line with psychological studies, my findings 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 estimation showed that bullying is very costly in terms of the 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 that the current stock of cognitive skills greatly influences 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 non-cognitive skills at early ages will greatly reduce victimization occurrence. My model also indicates that allocating students to more homogeneous classrooms might reduce victimization by preventing kids with uncommon characteristics from being isolated and targeted by bullies.

This paper intends to contribute to the human development literature in economics by exploring how school-aged kids' victimization 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 younger children's skill accumulation.

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., Bernal, R., Giannola, M., and Nores, M. (2020a). Child development in the early years: Parental investment and the changing dynamics of different dimensions. Working Paper 27812, National Bureau of Economic Research.

Attanasio, O., Bernal, R., Giannola, M., and Nores, M. (2020b). Child development in the early years: Parental investment and the changing dynamics of different dimensions. Working Paper 27812, National Bureau of Economic Research.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020c). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review*, 110(1):48–85.

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.*

Bjereld, Y., Daneback, K., and Petzold, M. (2017). Do bullied children have poor relationships with their parents and teachers? A cross-sectional study of Swedish children. *Children and Youth Services Review*, 73:347–351.

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.

Bowes, L., Maughan, B., Ball, H., Shakoor, S., Ouellet-Morin, I., Caspi, A., Moffitt, T. E., and Arseneault, L. (2013). Chronic bullying victimization across school transitions: The role of genetic and environmental influences. *Development and Psychopathology*, 25(2):333–346.

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.

Choi, H. and Choi, Á. (2015). When One Door Closes: The Impact of the Hagwon Curfew on the Consumption of Private Tutoring in the Republic of Korea.

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.

Del Bono, E., Kinsler, J., and Pavan, R. (2020). A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation. IZA Discussion Papers 13714, Institute of Labor Economics (IZA).

deLara, E. W. (2008). Developing a Philosophy about Bullying and Sexual Harassment: Cognitive Coping Strategies Among High School Students. *Journal of School Violence*, 7(4):72–96.

deLara, E. W. (2012). Why Adolescents Don't Disclose Incidents of Bullying and Harassment. *Journal of School Violence*, 11(4):288–305.

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*.

Freyberger, J. (2020). Normalizations and misspecification in skill formation models. Technical report, University of Bonn.

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. and Vytlacil, E. (2007). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of Econometrics*, pages 1–96.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2016). Dynamic treatment effects. *Journal of Econometrics*, 191(2):276–292.

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).

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

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.

Kochel, K. P., Ladd, G. W., and Rudolph, K. D. (2012). Longitudinal Associations Among Youth Depressive Symptoms, Peer Victimization, and Low Peer Acceptance: An Interpersonal Process Perspective. *Child development*, 83(2):637–650.

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.

Larranaga, E., Yubero, S., and Navarro, R. (2018). Parents' responses to coping with bullying: Variations by adolescents' self-reported victimization and parents' awareness of bullying involvement. *Social Sciences*, 7(8).

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.

NAS (2016). Preventing Bullying Through Science, Policy, and Practice. National Academies of Sciences, Engineering, and Medicine., Washington, DC.

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.

Nylund, K., Bellmore, A., Nishina, A., and Graham, S. (2007). Subtypes, Severity, and Structural Stability of Peer Victimization: What Does Latent Class Analysis Say? *Child development*, 78(6):1706–1722.

OECD (2012). *Education at a Glance 2012*. OECD Press, Paris.

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

OECD (2017). *PISA 2015 Results (Volume III) Students' Well-Being*. Students' Well-Being. OECD Publishing.

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.

Reijntjes, A., Kamphuis, J. H., Prinzie, P., and Telch, M. J. (2010). Peer victimization and internalizing problems in children: A meta-analysis of longitudinal studies. *Child Abuse & Neglect*, 34(4):244–252.

Ryoo, J. H., Wang, C., and Swearer, S. M. (2015). Examination of the change in latent statuses in bullying behaviors across time. *School Psychology Quarterly*, 30:105–122.

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

Scheithauer, H., Hayer, T., Petermann, F., and Jugert, G. (2006). Physical, verbal, and relational forms of bullying among German students: age trends, gender differences, and correlates. *Aggressive behavior*, 32(3):261–275.

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., 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., Shu, S., and Madsen, K. (2001). Characteristics of victims of school bullying developmental changes in coping strategies. In Graham, J. J. . S., editor, *Peer harassment in schools*, pages 332–351.

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.

Waasdorp, T. E. and Bradshaw, C. P. (2015). The Overlap Between Cyberbullying and Traditional Bullying. *Journal of Adolescent Health*, 56(5):483–488.

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

Wang, J., Iannotti, R., and Nansel, T. (2009). School Bullying Among Adolescents in the United States: Physical, Verbal, Relational, and Cyber. *Journal of Adolescent Health*, 45(4):368–375.

Appendix

A Attrition Analysis

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

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

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

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

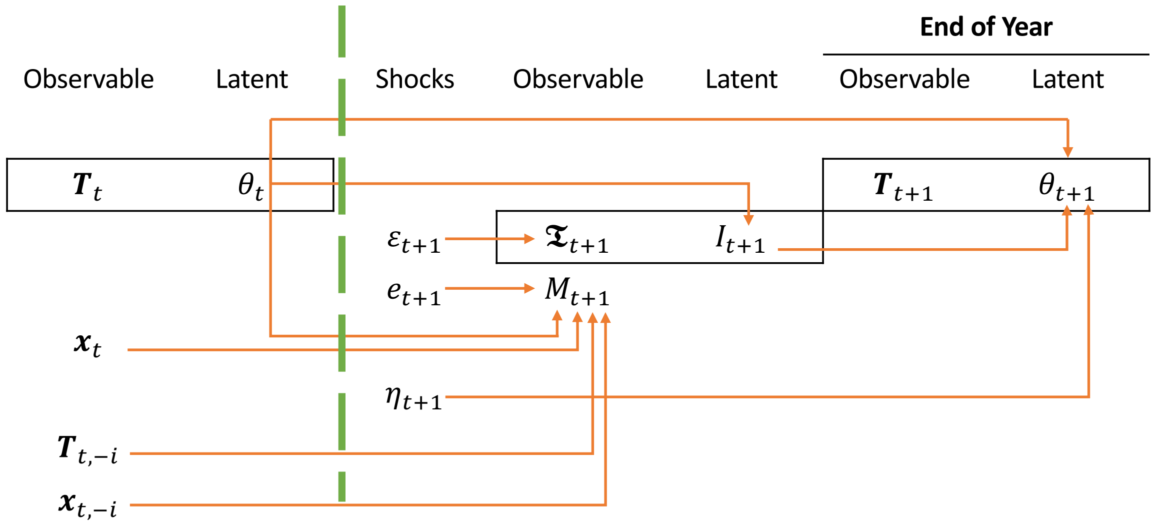
Table A.2: Probability of Staying from $t = 1$ to $t = 2$

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

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

B Model's Timeline

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



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

C Identification of the Model

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

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

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

and its off-diagonal elements are of the form:

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

where $\lambda_t^{T_\cdot,\cdot}$ are the elements of Λ_t^T. As it is, the measurement system is underidentified (Carneiro et al., 2003). Assumptions are needed. First, note that latent factors have no metric or scale of their own. This feature poses the need for normalizing to unity one loading per factor. Second, note that loadings, factor variances, and covariances need to be identified from the $L(L-1)/2$ off-diagonal elements of $COV\left(\mathbf{T}_t \,|\mathbf{X}_{t,T}\right)$ 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.[39] Note that this happens if $L \geq 6$.

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

$$\left[\begin{array}{c} \mathbf{T}_t^A \\ \mathbf{T}_t^{A,B} \end{array}\right] = \left[\begin{array}{c} \mathbf{X}_{t,T}\beta^T + \lambda_t^{(A)A}\theta_{A,t} + \mathbf{e}_t^{T^A} \\ \mathbf{X}_{t,T}\beta_t^T + \lambda_t^{(A,B)A}\theta_{A,t} + \lambda_t^{(A,B)B}\theta_{B,t} + \mathbf{e}_t^{T^{A,B}} \end{array}\right] \tag{13}$$

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

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

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

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

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

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

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

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

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

for $m, n = L_A+1, \ldots, L_B$, $m \neq n$ and $k = 1, \ldots, L_A$. It is easy to see that the second block of the measurement system is underidentified as it has $2L_B+1$ unknowns, while it has only ${}^{L_B(L_B+1)}/_2$ pieces of relevant information.[42] Therefore if $L_B = 3$, I have seven unknowns and six covariances to use for identification.

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

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

[42] Unknowns: two loadings per measure minus one that is normalized, $\sigma^2_{\theta^B}$ and $\sigma_{\theta^A,\theta^B}$. Measurement system covariances: ${}^{L_B(L_B-1)}/_2$ covariances within the second block measures and L_B covariances resulting from one covariance between each second block measure and one measure in the first block—preferably, the one that has the normalized loading.

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

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

$$COV\left(T_{t,m}^{A,B}, T_{t,o}^{A,B} \left|\mathbf{X}_T\right.\right) = \lambda_t^{(A,B)_m B} \sigma_{\theta_t^B}^2 + \lambda_t^{(A,B)_m A} \sigma_{\theta_t^A, \theta_t^B} \tag{15}$$

$$COV\left(T_{t,m}^{A,B}, T_{t,l}^{A} \left|\mathbf{X}_T\right.\right) = \lambda_t^{(A,B)_m A} \sigma_{\theta_t^A}^2 + \lambda_t^{(A,B)_m B} \sigma_{\theta_t^A, \theta_t^B} \tag{16}$$

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

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

D Identification of the CES Function

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

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

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

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

E Estimation of the Investment Factors

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

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

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