The Dynamic Consequences of Bullying on Skill Accumulation*†

Miguel Sarzosa‡
University of Maryland

January 18, 2015

Job Market Paper

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. I examine how bullying, an unfortunately common childhood experience, depletes skills in school children. I use a unique longitudinal data set on middle school students to estimate a structural dynamic model of skill accumulation based on the identification of unobserved heterogeneity. I allow skill formation to depend on past levels of skills, parental investment and bullying. Also, I allow bullying itself to depend on each student's past skills and those of his or her classmates. I find that being bullied at age 14 depletes current skill levels by 14% of a standard deviation for the average child. This skill depletion causes the individual to become 25% more likely to experience bullying again at age 15. Therefore bullying triggers a self-reinforcing mechanism that opens an ever-growing skill gap that reaches about one standard deviation by age 16. Finally, I find evidence that supports the allocation of students in more skill-homogeneous classrooms as a tool to reduce bullying occurrence.

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

JEL codes: I12, I14, I25, I31

---

*I would like to thank Sergio Urzua for his invaluable support. I would also like to thank John Ham, John Shea, Sebastián Galiani, Soohyung Lee, Brian Quistorff, the participants of the Global Education Forum at Seoul, Republic of Korea, and the participants of the microeconomics seminar at the University of Maryland, College Park, for their comments on earlier stages of this research project.

†For the latest version of this paper, please go to http://econweb.umd.edu/∼sarzosa/res/DynBullying.pdf.

‡3115P Tydings Hall, College Park, MD 20742-7211; phone: (301) 405-ECON (3266); fax: (301) 405-3542; email, sarzosa@econ.umd.edu. The author did not have access to information leading to the identification of individuals. The data analysis was carried out in a secure server.

1 Introduction

When Lim Jee-young, a South Korean woman from Daegu, read the suicide note left by her 13-year-old son, she found out that her young Seung-min had repeatedly been hit, robbed, burned and tortured by boys in his class.[1] Unfortunately, this horrible story is not unique. It repeats itself time and time again, not only in South Korea, but all around he world. The Columbine High School shooting in 1999 is still fresh in the collective memory of North Americans (US Department of Education, 2011),[2] as well as the story of Phoebe Prince, the 15-year-old girl that committed suicide in her house at South Hadley, Massachusetts, after having suffered several weeks of harassment and attacks while in school.[3] These and many other similar events—that exemplify the immense costs borne by bullying victims and communities in general—have increased the efforts made by societies to face and discuss bullying as behavioral issue among young people.[4]

Psychologists have defined a bullying victim as 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). 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. The existence of an imbalance of power is a key feature of bullying because this creates a sense of defenselessness of the victim. He or she may be outnumbered, physically weaker or less psychologically strong than the bullies (Smith and Brain, 2000). Faris and Felmlee (2011) suggest that bullying thrives in contexts where individuals need to show peer group status. 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

[1] http://www.cnn.com/2012/07/25/world/asia/south-korea-school-bully/

[2] A US Secret Service report shows that 2/3 of 37 school shootings involved attackers that felt "persecuted, bullied, threatened, attacked or injured by others" (Vossekuil et al., 2002).

[3] http://www.nytimes.com/2010/03/30/us/30bully.html?ref=us. This event induced the Massachusetts Senate to pass an anti-bullying law that carries her name.

[4] Anti-bullying campaigns and laws have been implemented in the US, Canada, UK, Germany, Scandinavia, Colombia and South Korea.

ground for bullying. Furthermore, cyberbullying has escalated this potential harm from bullying to new levels. The ability to instantly and widely convey information to peers facilitates harassment.

Bullying, even when not fatal, is very costly. According to the US government statistics reported on stopbullying.gov, 160,000 children miss school every day in the US because of fear of being bullied, representing 15% of those who do not show up to school;[5] one of every ten students drops out or changes school because of bullying; homicide perpetrators are twice as likely as homicide victims to have been bullied previously by their peers; and bullying victims are between 2 to 9 times more likely to consider suicide than non-victims. In the UK, at least half of suicides among young people are related to bullying. In South Korea, according to its Education Ministry, more than 77,000 students admitted to being bullied, and nearly 10 percent of those said they had considered suicide.

Surprisingly, economic literature has remained mostly silent on bullying. Very little is known about its intermediate costs and long-term consequences. In this paper, I contribute in bridging this gap by exploring the two-way relation between bullying and cognitive and non-cognitive skills accumulation. *Cognitive skills*, defined as "all forms of knowing and awareness such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving" (APA, 2006), and *non-cognitive skills*, vaguely defined as personality and motivational traits that determine the way individuals think, feel and behave (Borghans et al., 2008), are critical to the development of successful lives.[6] Hence, in this paper, I explore how school bullying hampers the development of successful adults by impeding optimal skill accumulation. In addition, I analyze how cognitive and non-cognitive skills are themselves important determinants of in-school victimization.

The proposed two-way relation between skills and bullying is based on facts about child victimization that have been widely established by the psychological literature. Psychologists have

[5] Kochenderfer and Ladd (1996) find a positive correlation between in-school victimization and school avoidance.

[6] See for example Murnane et al. (1995); Cawley et al. (2001); Heckman and Rubinstein (2001); Duckworth and Seligman (2005); Heckman et al. (2006a); Urzua (2008); Tough (2012); Prada and Urzua (2013)

noticed not only that bullying victims suffer grave and long-lasting consequences in terms of their emotional well being (Olweus, 1997; Smith and Brain, 2000, among many others), but also that the likelihood of being a victim increases dramatically when the child has some behavioral vulnerability (Hodges et al., 1997).[7]

To analyze this two-way relation, I develop a model that embeds peer-influenced skill contributions—like bullying—-in the skills accumulation process. I rely on the facts that skills are malleable and dynamic (Cunha and Heckman, 2007), and that many external conditions can influence the stock of skills an individual is endowed with at a given point in time (OECD, 2014). Little is known on how skill accumulation takes place, especially for non-cognitive skills. In particular, there is still a lot to be learned about the evolution of both cognitive and non-cognitive skills throughout childhood and teenage years, their dynamic interdependence, their two-way relationship with skill investment, the role social settings have on their development, and the long-term consequences of traumatic events and other skill-depleting forces like bullying. This paper intends to fill these gaps by quantifying the stock of skills lost to bullying by a given generation. The estimation strategy is based on the facts that skills beget skills (Cunha et al., 2010), that skills affect the occurrence of bullying (Sarzosa and Urzua, 2013a), and that bullying affects skill accumulation. Bullying victimization depletes current skill levels and not only lowers skill accumulation, but also makes an individual more prone to experience bullying again in the future, creating a self-reinforcing mechanism that generates a big burden to be carried during adulthood. This intuition is developed in a dynamic model of skill formation in which the bullying event is treated as a negative shock that depletes the existing stock of skills changing negatively the skill accumulation path for the people involved.

[7] In fact, Hodges et al. (1997) show that the occurrence of chronic victimization needs two conditions. First, the victim needs to display a behavioral vulnerability that may not only make the kid irritant to his or her peer, but also may signal that he or she might not be able to defend him or herself as they "cry easily, are anxious, lack humor, lack self confidence and self esteem". Second, victims occupy "a social position in the peer group that invites, disinhibits or permits aggressive attacks towards the child" .

Using a tractable model within the latent factor framework and extending the theoretical contributions of Cunha et al. (2010), I estimate the parameters that govern the process by which past skills determine future skills. I allow future skills to depend on current skills, current investment choices and whether one was bullied. I also allow for the investment choices, which I treat as a latent factor as well, to depend on the level of skills. In addition, I allow the bullying event to depend on both the personal skills endowments and the skills distributions of the peer group each student is exposed to. This setting allows me to estimate both the *direct* (i.e., how bullying at t affects future skills at $t+1$) and *indirect* (i.e., how bullying at t affects future skills at $t+2$ through changes in skills at $t+1$) channels through which bullying affects skill accumulation.

My structural model relies on the identification of latent skills to deal with selection. The model used for such identification is flexible enough to incorporate several desirable features. First, 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., 2006a). Second, the model uses a mixture of normals when estimating the distributions of the latent factors. This guarantees the flexibility required to appropriately recreate the unobserved distributions in the estimation. Therefore, it does not assume any functional form for the distributions of the factors; instead, I estimate them directly and often find that they are far from normal. Third, the model does not assume linearity in crucial steps of the estimation. Finally, the structural model allows me to simulate counterfactuals for each skill level, which I use to calculate the divergence in skill accumulation paths caused bullying.

This paper contributes to the economic literature in several ways. First, it inquires about the process of skill accumulation by providing a model that introduces peer characteristics into the dynamics of skill accumulation. Second, it analyzes the consequences of disruptive behavior in school in terms of skill depletion. Third, it extends my previous work on school bullying (Sarzosa and Urzua, 2013a), 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 to a generation. That is, I can go beyond school absenteeism and in-school stress, and estimate skill endowments loses for life. In addition, this will open an auspicious research agenda on skill accumulation and victimization in schools. The analysis could be extended in many directions. For instance, we could inquire the extent to which bullying can affect skill accumulation of non-direct victims, just because of the disruptiveness of the event.

This paper is organized as follows. Section 2 reviews the scarce related economic literature on the subject. In Section 3, I present the basic framework for the analysis of skill dynamics. Section 4 defines the empirical strategy I will use in this paper. Section 5 describes the data I use for the analysis, and describes of how the cognitive and non-cognitive skill measures are constructed. In section 6, I show my results and simulations. 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

This paper relates with two branches of the literature: bullying and skill formation.

2.1 Bullying

Economic research on bullying is scarce. Two main reasons explain this sparseness: first, lack of adequate data; and second, the non-randomness of the selection into bullying. Regarding the former, there is little longitudinal data that inquires about bullying, so there are very few sources that researches can use to observe individuals before and after the event. Regarding the latter, the non-randomness of bullying causes the consequences of bullying to be confounded by the intrinsic characteristics that made the person a victim or a perpetrator in the first place. While some studies have used longitudinal data, they have not been able to deal with the selection issue. To the best of my knowledge there are two papers in the economic literature that address bullying in particular. Brown and Taylor (2008) use OLS regressions and ordered probits to look at educational attainment and wages in the UK. They find that being bullied and being a bully are correlated with lower educational attainment and in consequence with lower wages later in life. Eriksen et al. (2012) use detailed Danish data, and estimate OLS and family-level FE regressions to find correlations between bullying and grades, pregnancy, use of psychopharmachological medication, height and weight. Although these are novel efforts, neither deal properly with the non-randomness of the bullying "treatment".[8]

The psychology and sociology literatures have been more prolific in examining bullying as a social phenomenon. For instance, Smith et al. (2004) show that bullying victims have fewer friends and are more likely to be absent from school. Psychological research surveyed by (Kim, 2005) show that victimized children are more likely to suffer bed wetting, sleep difficulties, anxiety, loneliness and isolation. Furthermore, Ouellet-Morin et al. (2011) show that victims' brains have unhealthy cortisol reactions that make it difficult to cope with stressful situations.

This literature has also found that younger kids are more likely to be bullied and that bullying is more frequent among boys than among girls (Boulton and Underwood, 1993; Perry et al., 1988; Dake et al., 2003). Interestingly, Olweus (1997) finds 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. However, he finds that victims are often smaller than attackers, and Lowenstein (1978) finds that victimized kids have more odd mannerisms than non-victimized kids.

The characterizations of the victims and perpetrators highlight the importance of controlling for

[8] Eriksen et al. (2012) argue that the family fixed-effects help address the selection on unobservables issue. However, I present evidence in Table 8 that shows that birth order influences the likelihood of being victimized at age 15.

non-cognitive skills throughout the analysis. According to psychological research, bullied children in general have less self-esteem, and have a negative view of their situation (Björkqvist et al., 1982; Olweus, 1997). They are also more likely to feel lonely (Dake et al., 2003). All these analyses, although descriptive as they are unable to find causal effects, provide a critical input in the definition of the models I use in my own work.

In Sarzosa and Urzua (2013a), we provide the first attempt to assess the determinants and consequences of bullying while dealing with the econometric problems caused by selection into becoming a victim or a perpetrator of bullying. We provide evidence distinguishing how cognitive and non-cognitive skills affect the likelihood of being bullied, being a bully, or being a cyberbully. We find that non-cognitive skills, but not cognitive skills, significantly reduce the chances of being bullied, being a bully and being a cyberbully during high school. Finally, in a non-dynamic context, we quantify the effect of being a bully and being bullied at age 15 on several outcomes at age 19 controlling for the unobserved heterogeneity caused by the latent skills. We find that both victims and bullies have negative consequences later in life. However, they differ in how non-cognitive and cognitive skills palliate or exacerbate these consequences.

In this paper, I contribute to the analysis of bullying literature by building on Sarzosa and Urzua (2013a) and providing an explanation for how the difference in the outcomes observed in that paper materialize. That is, while Sarzosa and Urzua (2013a) estimate the ATE of youth bullying on adult outcomes, in this present paper, I elucidate how these gaps are created, showing bullying as the triggering event that determines a divergence in skill accumulation paths.

2.2 Skill Dynamics

Although there has recently been a number of contributions in the area of skill development (Cunha et al., 2006, 2010), economic literature about that topic remains scarce. One of the main reasons for this scarcity is that it is difficult to directly measure cognitive and non-cognitive skills. At best, survey and administrative data contain measures that indirectly reflect underlying or latent skills. Structural models, like the one used in this paper, are needed to fully address this issue. Another difficulty arises due to the need of special longitudinal data sets that interview the subjects several times throughout particular periods of their life. Often this kind of data sets focus on labor, income and other economic dynamics and do not collect the measures needed to quantify skills.

What do we know so far? We know that skills are dynamic and malleable. That is, they depend on their past levels, they can be hindered and they can be fostered. Cunha et al. (2006) show that skills beget skills and therefore initial skill endowments and early accumulation are critical for the lifetime stock of skills. Based on this, they show that skill gaps between children from rich families and children from disadvantaged families start to widen at early ages. This gives foundation to the call for early childhood development and preschool interventions (Knudsen et al., 2006; Doyle et al., 2009). Skills beget skills not only though the natural process of getting the stock available at time t to $t+1$, but also though investment. There is some evidence on the fact that skills encourage skill investment. That is, skilled kids have higher levels of skill investment that lower skilled kids (Skinner and Belmont, 1993; Espinoza et al., 2014).[9]

The claim that skills are malleable is backed up by a series of papers that show that skill developing interventions were able to modify the stock of skills of the treated population. For instance, Heckman et al. (2010) show that the people treated by Perry Preschool Program have higher non-cognitive skills –although similar levels of cognitive skills– than the controls. 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). Cunha et al. (2010) show that skill developing interventions can compensate for low initial level of both cognitive and non-cognitive skills. There is evidence, however, that there are windows

[9]Although this feature of skill dynamics is widely proposed in Cunha et al. (2010), their theoretical claim of skills inducing higher levels of investment is only backed up by their empirical estimates in very early stages of life (i.e., before two years of age).

of opportunity outside of which skill malleability is lost (Knudsen et al., 2006). Cunha et al. (2006) argue that such window closes earlier for cognitive than for non-cognitive skills.

Besides the dynamism and malleability features of skills, recent literature has found that they strongly depend on different contexts the child grows in.[10] For instance, extensive literature finds that family background influences skill accumulation: children whose parents are more engaged in their upbringing are likely to have higher levels of both types of skill.[11] The quality of school inputs such as class size and teacher characteristics also affects non-cognitive skills (Fredriksson et al., 2013; Jackson, 2013). Skill endowments have even been found to depend on the level of stress a person was exposed to during childhood (McEwen and Seeman, 2006).

3 Skill Formation With Peer-Influenced Investment

Economic and psychological literature has found skills to be dynamic and prone to be influenced by many external conditions. Hence, my framework needs to incorporate five facts that result from this 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 can hamper skill development, and v. bullying 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. Let the stock of skills $S = \{A, B\}$ a person i that belongs to classroom c has at time $t+1$ (i.e., $\theta_{S,i\in c,t+1}$) be a result of a CES skill production function whose inputs are the stock of skills she had at time t ($\theta_{A,i\in c,t}$ and $\theta_{B,i\in c,t}$), the skill investment choices done between the two periods ($I_{S,i\in c,t}$), and the occurrence of a skill depleting shock like bullying ($M_{i\in c,t}$). Furthermore, I allow for the investment

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

[11] See Hart and Risley (1995); Cunha et al. (2006); Heckman and Masterov (2007); Cabrera et al. (2007); Kiernan and Huerta (2008); Tough (2012)

choice and the of bullying occurrence to be affected by the previous levels of skills.

$$\theta_{S,i\in c,t+1} = \left[\gamma_{A,t}\theta^{\rho}_{A,i\in c,t} + \gamma_{B,t}\theta^{\rho}_{B,i\in c,t} + \gamma_{I,t}I^{\rho}_{S,i\in c,t+1} + \gamma_{M,t}M^{\rho}_{i\in c,t+1}\right]^{1/\rho} \tag{1}$$
$$I_{S,i\in c,t+1} = \alpha^{S}_{A,t}\theta_{A,i\in c,t} + \alpha^{S}_{B,t}\theta_{B,i\in c,t} + \nu_{t+1}$$
$$M_{i\in c,t+1} = \mathfrak{h}\left(\theta_{A,i\in c,t}, \theta_{B,i\in c,t}, \theta_{A,-i\in c,t}\right)$$

for $S = \{A, B\}$, where $\gamma_{M,t} = 1 - \gamma_{A,t} - \gamma_{B,t} - \gamma_{I,t}$ and $-i \in c$ indicates all individuals that belong to classroom c except i. Let N_c be the total number of students in classroom c, where $c \in C$.

This structure relates parental investment choices with victimization at $t+1$ indirectly through its effect on the stock of skills at t. This relies on the results of psychological research that indicates that responsive and supporting parenting practices are related with lower levels of bullying (Flouri and Buchanan, 2002). In particular, certain parental behaviors that hamper the development of locus of control on kids have been linked with in-school victimization (Ladd and Ladd, 1998).

In the system of equations (1), $\mathfrak{h}(\cdot)$ is a function that maps $\mathbb{R}^{N_c+1} \to \mathbb{R}$ representing a sufficient statistic that relates the skills available to other students in the classroom c with those of individual $i \in c$ that affects the likelihood of i being bullied. Through function $\mathfrak{h}(\cdot)$, I incorporate two stylized facts of bullying established by the psychological literature: i. that there are personal characteristics of the student that influences the chances of being bullied (i.e., behavioral issues), and ii. that there are characteristics of the peer group that set him or her apart from his or her classmates (e.g., lacks friends, is rejected by the peer-group) (Hodges et al., 1997). Function $\mathfrak{h}(\cdot)$ responds to the fact that bullying needs a social arena in which the imbalances of power take place allowing classmates to play different roles: victim, perpetrator and bystanders.[12] Therefore, the question that arises is: what separates bystanders from victims. Due to its social setting, one may

[12]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.

be inclined to look for answers to this question in the social interactions literature as in Schelling (1971), Pollak (1976) and Manski (1993), where agents interact through their decisions. The problem with bullying is that no one *decides* to be a victim. Hence, while the social interactions literature explains "why do members of the same group tend to behave similarly" (Manski, 2000), we are instead interested in answering "why is this kid chosen among the rest". Hence, "selection into bullying" is non-random and, like in a *social interactions* situation, it depends on characteristics of the victim and its classmates, but in a very different way. The idea is that individual i with skills set $(\theta_{A,i,t}, \theta_{B,i,t})$ might be bullied in classroom c but not in classroom c'. This difference depends on the skill distribution of the other students that belong to each classroom. This defines a different dimension of peer-influenced consequences. Therefore, identification relies on the assumption that the allocation of individual i to classroom c was exogenous, and therefore the assignment of i's classmates is as good as random.[13]

The choice of the CES as the production function of skills responds to two main reasons. First, its capability to smoothly introduce investment and bullying as inputs together with past levels of skill. Second, it provides the curvature needed to explore complementarieties between the inputs involved in the skill production function. In particular, the *static complementariety* $(\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial\theta_{S,t})$ and the *dynamic complementariety* $(\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial I_{S,t})$, concepts introduced by Cunha and Heckman (2008) to describe how the current stock of skills affect the productivity of skill investment, and how much of that productivity of investment is leveraged by past investment choices. I will use the same concepts to analyze the *skill depleting* power of the bullying event. Therefore the static complementariety on bullying will be given by

[13] This framework is particularly relevant in the South Korean context where after 1969 a "leveling policy" was introduced to regulate student placement. According to Kang (2007) "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." The leveling policy also makes the grouping of students by ability and achievement levels "extremely rare". These features let Kang (2007) claim that "the non-grouping (or ability mixing) in school exposes students to a classroom peer group that is nearly exogenously and randomly determined." 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.

$$\frac{\partial^2 \theta_{S,t+1}}{\partial M_{t+1} \partial \theta_{S,t}} \tag{2}$$

and the dynamic complementariety of bullying by

$$\frac{\partial^2 \theta_{S,t+1}}{\partial M_{t+1} \partial M_t} \tag{3}$$

4 Empirical Strategy

The key feature of the empirical strategy is the way it deals with the fact that underlying cognitive and non-cognitive skills and investment preferences are latent rather than observable.[14] 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).

In this chapter, first I present how I use manifest scores to identify the latent variables of interest and then, based on that, I show how this allows me to estimate the full model.

4.1 Identification of Latent Factors' Distributions: From the Static to the Dynamic Setting

The core of the empirical strategy is the assumption of a linear relation between the manifest and the latent variables, that can be thought of as a production function of scores, whose inputs include both the individual observable characteristics and the latent endowments. In that sense, the empirical strategy incorporates the fact that the observed manifest values respond not only to

[14]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.

the latent variables of interest (θ), but also to observable traits ($\mathbf{X}$) and random shocks (e^T) in the following form:

$$\mathbf{T}_t = \mathbf{X}_{t,T}\beta_t^T + \alpha_{\mathbf{t}}^{\mathbf{T,A}}\theta_t^A + \alpha_{\mathbf{t}}{}^{\mathbf{T,B}}\theta_t^B + \mathbf{e}_{\mathbf{t}}^{\mathbf{T}} \tag{4}$$

where $\mathbf{T}_t$ is a $L \times 1$ vector of measurements (e.g., test scores) at time t, $\mathbf{X}_{t,T}$ is a matrix with all observable controls for each measurement at time t and $\alpha_{\mathbf{t}}{}^{\mathbf{T,A}}$ and $\alpha_{\mathbf{t}}^{\mathbf{T,B}}$, are the loadings of the unobserved factors at time t. I assume that $\left(\theta_{\mathbf{t}}{}^{\mathbf{A}}, \theta_{\mathbf{t}}{}^{\mathbf{B}}, \mathbf{X}_{t,T}\right) \perp \mathbf{e}_{\mathbf{t}}^{\mathbf{T}}$, that all the elements of the $L \times 1$ vector $\mathbf{e}_{\mathbf{t}}^{\mathbf{T}}$ are mutually independent and have associated distributions $f_{e_t^h}(\cdot)$ for every $h = 1, \ldots, L$.[15]

Carneiro et al. (2003), based on the insights of Kotlarski (1967), show that identification of the loadings in (4) (up to one normalization[16]) and the (diagonal) matrix of the variances of the latent factors Σ_θ needs three restrictions:

R1 Orthogonality of the factors (i.e., $\theta^A \perp \theta^B$).

R2 L to be at least $2k + 1$, where k is the number of latent factors in the model.[17]

R3 The factor structure within the measurement system (4) needs to follow a triangular pattern

[15]Identification of latent investment preferences in skill S, I_t^S, follows the same structure:

$$\mathcal{I}_{S,t} = \mathbf{X}_{t,\mathcal{I}}\beta_t^{\mathcal{I}_S} + \alpha_{\mathbf{t}}^{\mathcal{I}_{\mathbf{S}}} I_{A,t} + \mathbf{e}_{\mathbf{t}}^{\mathcal{I}_{\mathbf{S}}}$$

where $\mathcal{I}_{S,t}$ is a vector with the investment manifest scores.

[16]That is, one of the loadings $\alpha_t^{T,\cdot}$ of each factor should be set equal to 1, and the estimation of all the rest of the loadings should be interpreted as relative to those used as nummeraire.

[17]For simplicity, I will assume that I have $3k$ measurements in a measurement system like (4).

like

$$\begin{bmatrix} \alpha^{T,A} & \alpha^{T,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & \alpha^{T_1,B} \\ \alpha^{T_2,A} & \alpha^{T_2,B} \\ \alpha^{T_3,A} & \alpha^{T_3,B} \\ \alpha^{T_4,A} & \alpha^{T_4,B} \\ \alpha^{T_5,A} & \alpha^{T_5,B} \\ \alpha^{T_6,A} & \alpha^{T_6,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & 0 \\ \alpha^{T_2,A} & 0 \\ 1 & 0 \\ \alpha^{T_4,A} & \alpha^{T_4,B} \\ \alpha^{T_5,A} & \alpha^{T_5,B} \\ \alpha^{T_6,A} & 1 \end{bmatrix} \tag{5}$$

which indicates that the first three manifest scores are affected by the first factor only, while the second three manifest scores are affected by both latent factors.[18]

However, these restrictions only apply to cases where there are no factor dynamics involved in the estimation. It is easy to see that Restriction R1 cannot be sustained if we believe there are dynamics governing the production of factor endowments at a given point in time. In particular, in a dynamic and intertwined process in which $\theta_{S,t+1} = g_S\left(\theta_t^A, \theta_t^B\right)$ for $S = \{A, B\}$, $\theta_{t+1}^A \not\perp \theta_{t+1}^B$ holds because of common past influences. That is, θ_{t+1}^A and θ_{t+1}^B are correlated because both share common inputs θ_t^A and θ_t^B, even if each latent factor has its own production function $g_A(\cdot,\cdot)$ and $g_B(\cdot,\cdot)$.

In order to get rid of the orthogonality assumption of contemporaneous latent factors R1 and still be able to identify the latent factors' distributions and loadings from a measurement system like (4), I need to rely on a factor structure different from the one required in Restriction R3. I

[18]The loading structure of (4) 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 (4) we will have the simplest version of (5):

$$\begin{bmatrix} \alpha^{T,A} & \alpha^{T,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & \alpha^{T_1,B} \\ \alpha^{T_2,A} & \alpha^{T_2,B} \\ \alpha^{T_3,A} & \alpha^{T_3,B} \\ \alpha^{T_4,A} & \alpha^{T_4,B} \\ \alpha^{T_5,A} & \alpha^{T_5,B} \\ \alpha^{T_6,A} & \alpha^{T_6,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & 0 \\ \alpha^{T_2,A} & 0 \\ 1 & 0 \\ 0 & \alpha^{T_4,B} \\ 0 & \alpha^{T_5,B} \\ 0 & 1 \end{bmatrix} \tag{6}$$

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).

need to assume a factor structure (5) where $\alpha^{T_6,A} = 0$ (i.e., for each latent factor, there is at least one manifest score that is only affected by that factor).

Theorem 1. *If $\theta_{t+1}^{A} \not\perp \theta_{t+1}^{B}$ and there is, at least one test score per latent factor that depends only on one latent factor, then the factor loadings of a measurement system like (4) are identified.*

Proof. See Appendix B.1 □

Contrary to common practice latent factor literature, I do not impose normality to the distribution of the factors $f_{\theta^A,\theta^B}(\cdot,\cdot)$. Instead, I use the mixture of normals in order to achieve the flexibility required to mimic the true underlying distributions of the latent endowments. The mixture of normals not only grants flexibility in the type of distribution it is able to replicate, but also allows numerical integration using the Gauss-Hermite quadrature, which is particularly useful for calculating $E[f(X)]$ when $X \sim \mathcal{N}(\mu, \sigma^2)$ (Judd, 1998).[19] Therefore, the likelihood is

$$\mathcal{L} = \prod_{i=1}^{N} \int \int \left[\; f_{e_t^1}\left(\mathbf{X}_{t,T_1}, T_{t,1}, \zeta^A, \zeta^B\right) \times \cdots \times f_{e_t^L}\left(\mathbf{X}_{t,T_L}, T_{t,L}, \zeta^A, \zeta^B\right) \; \right] \Delta F_{\theta_t^A,\theta_t^B}\left(\zeta^A, \zeta^B\right) \tag{7}$$

where I integrate over the distributions of the factors due to their unobservable nature, obtaining $\hat{\beta}_t^T$, $\hat{\alpha}_t^{T,A}$, $\hat{\alpha}_t^{T,B}$, $\hat{F}_{\theta_t^A}(\cdot)$ and $\hat{F}_{\theta_t^B}(\cdot)$.

Using this latent variable framework allows me to use a construct that lacks metric and measuring instruments and disentangle the variation of interest (i.e., that one that comes from the unobserved heterogeneity) from the one generated by random shocks and the one that comes from exogenous observable traits like gender or age. In fact, variance decompositions of the test scores used for the estimation and presented in Figure 1 show that latent endowments explain between 5 to 9 times more the variation of the scores than the observable characteristics. However, Figure 1 also shows that even after controlling for latent endowments more that half of the variation of

[19]The structural estimations presented in this paper were done using the `heterofactor` command for Stata developed by Miguel Sarzosa and Sergio Urzua (Sarzosa and Urzua, 2012).

the scores is still unexplained. Hence, these findings are in line with the argument against the use test scores as proxies of abilities.

4.2 Dynamic Estimation

4.2.1 Identification and Estimation Steps

As explained above, the core of the empirical strategy is that $\theta_{S,t}$ and $I_{S,t}$ $\forall S, t$ are unobservable factors. Therefore, we need to estimate all the parameters of equations (4), and of the dynamic model (1) using maximum likelihood estimation procedures by integrating over the distributions that describe the latent factors. In this section, I will describe the steps and the identification sources involved in the estimation of the dynamic process described in (1).

Suppose the data we use follows individuals for two time periods: t and $t+1$. Although in the actual estimations I use a triangular loading matrix like (5) with $\alpha^{T_6,A} = 0$, for simplicity suppose the manifest cognitive and non-cognitive scores have a loading structure presented in equation (6). Therefore the measurement system is the following:

$$\mathbf{T}_{A,t} = \mathbf{X}_{t,T}\beta_t^{T_A} + \alpha_{\mathbf{t}}^{\mathbf{T_A}}\theta_t^A + \mathbf{e}_{\mathbf{t}}^{\mathbf{T_A}} \tag{8}$$

$$\mathbf{T}_{B,t} = \mathbf{X}_{t,T}\beta_t^{T_B} + \alpha_{\mathbf{t}}^{\mathbf{T_B}}\theta_t^B + \mathbf{e}_{\mathbf{t}}^{\mathbf{T_B}} \tag{9}$$

$$\mathbf{T}_{A,t+1} = \mathbf{X}_{t+1,T}\beta_{t+1}^{T_A} + \alpha_{\mathbf{t+1}}^{\mathbf{T_A}}\theta_{t+1}^A + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_A}} \tag{10}$$

$$\mathbf{T}_{B,t+1} = \mathbf{X}_{t+1,T}\beta_{t+1}^{T_B} + \alpha_{\mathbf{t+1}}^{\mathbf{T_B}}\theta_{t+1}^B + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_B}} \tag{11}$$

$$\mathcal{I}_{A,t+1} = \mathbf{X}_{t+1,\mathcal{I}}\beta_{t+1}^{\mathcal{I}_A} + \alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{A}} I_{A,t+1} + \mathbf{e}_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{A}} \tag{12}$$

$$\mathcal{I}_{B,t+1} = \mathbf{X}_{t+1,\mathcal{I}}\beta_{t+1}^{\mathcal{I}_B} + \alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{B}} I_{B,t+1} + \mathbf{e}_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{B}} \tag{13}$$

$$M_{t+1} = \mathfrak{h}\left(\mathbf{X}_{t+1,M}\beta_{t+1}^M, \alpha_{\mathbf{t+1}}^{\mathbf{M}}\theta_{i\in c,t}, \alpha_{\mathbf{t+1}}^{\mathbf{M_c}}\theta_{-i\in c,t}\right) + e_{t+1}^M \tag{14}$$

where $\mathbf{T}_{S,\tau}$ is a 3×1 vector that contains each of the test scores associated to skill $S = \{A, B\}$

at time $\tau = \{t, t+1\}$, and $\mathcal{I}_{S,t+1}$ is a 3×1 vector that contains each of the investment measures made in skill $S = \{A, B\}$ at time $t+1$. As shown in the previous section, we can use equations (8) and (9) to identify $\hat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$, and equations (12) and (13) to identify $\hat{F}_{I_{S,t+1}}(\cdot)$. Also, we can use (10) and (11) to consistently estimate $\hat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$ and $\hat{\beta}^{T_S}_{t+1}$. In consequence, I are able to construct the vectors

$$\xi_{A,t+1} = \mathbf{T}_{A,t+1} - \mathbf{X}_{t+1,T}\hat{\beta}^{T_A}_{t+1} = \alpha^{\mathbf{T_A}}_{\mathbf{t+1}}\theta^{A}_{t+1} + \mathbf{e}^{\mathbf{T_A}}_{\mathbf{t+1}} \tag{15}$$

$$\xi_{B,t+1} = \mathbf{T}_{B,t+1} - \mathbf{X}_{t+1,T}\hat{\beta}^{T_B}_{t+1} = \alpha^{\mathbf{T_B}}_{\mathbf{t+1}}\theta^{B}_{t+1} + \mathbf{e}^{\mathbf{T_B}}_{\mathbf{t+1}} \tag{16}$$

We now substitute the CES production function from (1) in the 3×1 measurement system for $\xi_{S,t+1}$. For instance, in the case when $S = A$ we have

$$\xi_{A,t+1} = \alpha^{\mathbf{T_A}}_{\mathbf{t+1}}\left[\gamma_{A,t}\theta^{\rho}_{A,i\in c,t} + \gamma_{B,t}\theta^{\rho}_{B,i\in c,t} + \gamma_{I,t}I^{\rho}_{S,i\in c,t+1} + \gamma_{M,t}M^{\rho}_{i\in c,t+1}\right]^{1/\rho} + \mathbf{e}^{\mathbf{T_A}}_{\mathbf{t+1}}$$

where $\gamma_{M,t} = 1 - \gamma_{A,t} - \gamma_{B,t} - \gamma_{I,t}$, from which I can identify $\gamma_{A,t}$, $\gamma_{B,t}$, $\gamma_{I,t}$ and ρ through a ML estimation assuming additive separability of the error term $\mathbf{e}^{\mathbf{T_S}}_{\mathbf{t+1}}$ and taking advantage of the fact that I have already identified $\alpha^{T_S}_{t+1}$. Hence, the likelihood function in the case where $S = A$ is:

$$\begin{aligned}
\mathcal{L} = \prod_{i=1}^{N}\int\int\int \; & f_{e^{T_{1,A}}_{t+1}}\left(\xi^{1}_{A,t+1} - \alpha^{T_{1,A}}_{t+1}\left[\gamma_{A,t}\theta^{\rho}_{A,i\in c,t} + \gamma_{B,t}\theta^{\rho}_{B,i\in c,t} + \gamma_{I,t}I^{\rho}_{A,i\in c,t+1} + \gamma_{M,t}M^{\rho}_{i\in c,t+1}\right]^{1/\rho}\right) \\
& \times f_{e^{T_{2,A}}_{t+1}}\left(\xi^{2}_{A,t+1} - \alpha^{T_{2,A}}_{t+1}\left[\gamma_{A,t}\theta^{\rho}_{A,i\in c,t} + \gamma_{B,t}\theta^{\rho}_{B,i\in c,t} + \gamma_{I,t}I^{\rho}_{A,i\in c,t+1} + \gamma_{M,t}M^{\rho}_{i\in c,t+1}\right]^{1/\rho}\right) \\
& \times f_{e^{T_{3,A}}_{t+1}}\left(\xi^{3}_{A,t+1} - \left[\gamma_{A,t}\theta^{\rho}_{A,i\in c,t} + \gamma_{B,t}\theta^{\rho}_{B,i\in c,t} + \gamma_{I,t}I^{\rho}_{A,i\in c,t+1} + \gamma_{M,t}M^{\rho}_{i\in c,t+1}\right]^{1/\rho}\right) \\
& \times f_{\nu_{t+1}}\left(I^{\rho}_{A,i\in c,t+1} - \alpha^{A}_{A,t}\theta_{A,i\in c,t} - \alpha^{A}_{B,t}\theta_{B,i\in c,t}\right) \\
& \times f_{e^{M}_{t+1}}\left(M_{t+1} - \mathfrak{h}\left(\mathbf{X}_{t+1,M}\beta^{M}_{t+1}, \alpha^{\mathbf{M}}_{\mathbf{t+1}}\theta_{i\in c,t}, \alpha^{\mathbf{M_c}}_{\mathbf{t+1}}\theta_{-i\in c,t}\right)\right)\Delta F_{\theta^{A}_{t},\theta^{B}_{t}}\left(\zeta^{A},\zeta^{B}\right)dF_{I_{S,t+1}}\left(\zeta^{I}\right)
\end{aligned}$$

Note that given that I allow $\mathbf{X}_{t,T}$ and $\mathbf{X}_{t+1,T}$ to have a constant term, the estimated factors are centered at zero. That is, overall skill means are absorbed in the constant terms of the $\beta_{\tau}^{T_{3},S}$ vectors for $\tau = \{t, t+1\}$ in equations (8) through (11). Estimation relies on that fact to identify ρ as shown in Theorem 2.

Theorem 2. *If $E[\theta_{A,t}] = 0$, $E[\theta_{B,t}] = 0$, the error e is additive separable in $\theta_{t+1} = f(\theta_t, \gamma, \rho) + e$, and $f(\theta_t, \gamma, \rho)$ is double differentiable , the parameter ρ is identified out of the fact that $E[\theta_{\cdot,t+1}] = 0$*

Proof. See Appendix B.2 □

Theorem 2 has an important implication regarding what I am able to capture in the dynamic process. The parameters estimated from (1) exclude effects due to the overall mean changes in skills. Suppose for instance, that everyone's cognitive skills increase on average by ϵ from t to $t+1$, such overall improvement does not affect the estimates of $\gamma_{A,t}$, $\gamma_{B,t}$, $\gamma_{I,t}$ and ρ . That is, my dynamic process will only capture idiosyncratic changes along the distributions. However, I can rely in the findings of Urzua (2008) who shows that under mild linearity assumptions in the measurement systems (8) through (11), I can claim that the mean of the skills is given by the coefficients in $\beta_{\tau}^{T_{3},S}$ associated with the constant term, call them $\beta_{\tau}^{T_{3},S}[1]$ for $\tau = \{t, t+1\}$. Therefore, I can retrieve the overall mean changes of skill S between t and $t+1$ (i.e., $\Delta_{t+1,t}^{S}$) by subtracting these constants. That is, $\Delta_{t+1,t}^{S} = \beta_{t+1}^{T_{3},S}[1] - \beta_{t}^{T_{3},S}[1]$.

4.2.2 The Choice of $\mathfrak{h}(\cdot)$ and the Identification of the Bullying Equation

As explained above, psychologists have established that in-school victimization occurrence requires some intrinsic characteristics of the person herself and characteristics of the person *vis-a-vis* the peer group (Hodges et al., 1997). The personal intrinsic traits are introduced in function $\mathfrak{h}(\cdot)$ through the observable and unobservable characteristics. The characteristics within the group, that is, the characteristics that "invite, disinhibits, or permits" attacks towards a given child are modeled by how uncommon in terms of particular traits the potential victim is. This uncommonness feature is important because it relates with several established facts in the psychological literature. First, that there needs to exist and imbalance of power (Olweus, 1997; Smith and Brain, 2000), therefore a kid with uncommon characteristics is less likely to have friends who can defend him or her (Bukowski et al., 1995). Also, bullies are more likely to attack those with no friends (Hodges and Perry, 1996). Furthermore, kids with uncommon characteristics are more easily regarded as weird and unlikeable, which fosters peer rejection (Hodges et al., 1997).[20]

The measure of rarity or uncommonness is materialized in my model by the number of classmates that lie inside an epsilon-ball in the skills or 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, you have higher chances of being bullied. So, if $\nabla_{\psi,i\in c}(d)$ is the number of classmates of i that lie in an epsilon-ball with radius d in the space of characteristic ψ, then the $\mathfrak{h}(\cdot)$ function becomes

$$M_{i\in c,t+1} = \mathfrak{h}\left(\mathbf{X}_M, (\theta_A, \theta_B)_{i\in c,t}, \nabla_{\psi,i\in c}(d), \mu_c\right) \tag{17}$$

Where μ_c is a school fixed effect that responds to the fact that there are several school characteristics like teachers quality, or overall faculty tolerance to bullying that influence the likelihood of bullying victimization (Dake et al., 2003). In addition, I incorporate school district fixed-effects to properly take advantage of the school allocation randomness in the South Korean education system.

This way to introduce classmates' characteristics is also econometrically advantageous as it goes around the well known problem of peer-effect identification. According to Angrist (2014) randomness in peer allocation is not sufficient to identify peer-effects. He claims that, in order to

[20] 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.

prevent the unwanted existence of mechanical statistical forces that create spurious correlations, the econometrician needs that not everyone within the group becomes affected or "treated" by the same peer-effect.[21] In my approach, the uncommonness measure allows for a different "treatment" for every observation to the point that, although everyone is affected by what happens inside their particular epsilon-ball, the relative position of those classmates that do not fall within that epsilon-ball is irrelevant.

In the estimations presented in this paper I consider bullying to be a dichotomous variable that takes the value of 1 if the person i was bullied and 0 if not. Hence, in this particular case, (17) becomes

$$M_{t+1} = \Phi\left(\mathbf{X}_{M,i\in c}\beta^{M} + \alpha_{t+1}^{M_A}\theta_{A,i\in c,t} + \alpha_{t+1}^{M_B}\theta_{B,i\in c,t} + \alpha_{t+1}^{\nabla_{\theta_S}(d)}\nabla_{\psi,i\in c}(d) + \mu_c + e_{t+1}^{M} > 0\right)$$

where $\Phi(\cdot)$ is the normal CDF.

5 Data

I empirically estimate the described model using the Junior High School Panel (JHSP) of the Korean Youth Panel Survey (KYP). I choose to use these data motivated by two reasons. First, bullying is a very important social issue in the South Korean society, probably more so than anywhere else in the world as they have an active policy aimed to curve the incidence of suicide. Suicide figures in South Korea are striking. It causes 31.7 deaths per 100,000 people, the single highest rate in the world. 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. Suicide is often linked with school bullying. Statistics of the South Korean

[21] For instance, measuring peer-effects by introducing a classroom mean is invalid as everyone is being "treated" by that classroom mean which will create a tautological relation captured in the regression by the coefficient associated with the mean.

Education Ministry show that more than 77,000 students admitted to being bullied, and nearly 10 percent of those said they had considered suicide. In response to this, bullying has recently been placed at the center of social policy in South Korea by president Park Geun-hye as it has been considered to be one of the "four social evils" together with sexual assault, domestic violence and food contamination.[22] Since 2012, the government installed more than 100,000 closed-circuit cameras in school facilities to prevent bullying and prosecute its perpetrators.

The second reason to focus on South Korea is data availability and KYP-JHSP's unique sampling scheme that allows the identification of peer characteristics. The KYP-JHSP is a longitudinal survey that started in 2003 sampling full junior high-school classrooms (i.e., 14 year olds) from which all the students were interviewed. They were interviewed once a year until 2008. Thus, they were followed though high-school and into the beginning of their adult life.

As this is a sensitive age range regarding life-path choices, the KYP-JHSP provides an interesting opportunity to understand the effects of non-cognitive skills on later decisions and behavior. The KYP-JHSP pays special attention to the life-path choices made by the surveyed population, inquiring not only about their decisions, but also about the environment surrounding their choices. Surveyed youths are often asked about their motives and the reasons that drive their decisions. Future goals and parental involvement in such choices are frequently elicited. The KYP-JHSP is also suitable to track non-cognitive skill dynamics given that the kids are interviewed for the first time at the beginning of their teen period. This allows the researcher to observe the evolution of skills during this critical age, and to see how the quality of the teenager's environment affects the likelihood of making good choices and avoiding risky and harmful behavior.

The sample consists of 12 regions including Seoul Metropolitan City. Very importantly for this research project, entire classrooms were sampled according to the proportion of national second year junior high-school students present in each region. All of the students of the sampled

[22] http://www.bbc.com/news/world-asia-26080052

classrooms were interviewed. The panel consists of 3,449 youths and their parents or guardians (see descriptive statistics in Table 1). Subjects were consistently interviewed in six waves.[23] Each year, information was collected in two separate questionnaires: one for the teenager, and another one for the parents or guardians.

Besides inquiring about career planning and choices, the KYP-JHSP inquires about academic performance, student effort and participation in different kinds of private tutoring. The survey also asks about time allocation, leisure activities, social relations, attachment to friends and family, participation in deviant activity, and victimization in different settings including bullying.[24] In addition, the survey asks a comprehensive battery of personality questions from which measures of self-esteem, self-stigmatization, self-reliance, aggressiveness, anger and self control can be constructed.

While the youths are often asked about the involvement of their parents in many aspects of their life, parents and guardians answer only a short questionnaire covering household composition and their education, occupation and income.

5.1 Reported Bullying

Bullying, as all other personal characteristic that was collected in the KYP-JHSP, is self-reported by the students. It refers to events where they have been severely teased or bantered, threatened, collectively harassed, severely beaten, or robbed. Hence, even though psychologists have constructed a very wide definition for bullying which I presented in the introduction of this paper, the kids in the study respond to the most direct and less subtle versions of bullying. This is in line with the findings in several international studies (see Madsen, 1996; Smith and Levan, 1995;

[23] 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.

[24] 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.

Smith et al., 1999, 2002) where children have been found to "focus on the more obvious and less subtle forms of bullying such as direct verbal and physical abuse and overlook indirect aggression" (Naylor et al., 2010). In the same way, the reported incidence of bullying in the KYP-JHSP, presented in Table 2, is in line with the incidence reported in international studies (see Smith and Brain, 2000, for a summary).

5.2 The Construction of the Non-Cognitive Skill Measures

As mentioned below in the description of the empirical strategy, the estimation of the distribution parameters of the latent non-cognitive trait uses three scores that measure socio-emotional skills. The KYP-JHSP contains a comprehensive battery of measures related to socio-emotional skills. Among them, I use the measures of locus of control, responsibility and self-esteem in the initial estimation of the distribution of non-cognitive skills.

It should be noted that most of the socio-emotional information in the KYP-JHSP is recorded in categories that group the reactions of the respondent in bins like "strongly agree" or "disagree". In consequence, and following common practice in the literature, I construct socio-emotional skill measures by adding categorical answers of several questions regarding the same topic. This method incorporates some degree of smoothness in the scores, which is essential for the estimation procedure. The questions used can be found in Appendix C.

5.3 The Construction of the Cognitive Skill Measures

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

6 Results

6.1 Skill Distributions

Tables 3 and 4 show the results of the estimation on the measurement system used to identify the joint skill distribution for $t = 1$ and $t = 2$, respectively. These distributions are presented in Figure 2a and 2b. They show that skill distributions are far from normal, and that there is a positive correlation between both dimensions of skills. In fact it is estimated to be of about 0.3869 for $t = 1$ and 0.358 for $t = 2$. This indicates that kids with high levels of one skill tend to have high levels of the other skills as well. An additional interesting feature of the joint skills distribution is the fact that the variance of non-cognitive skills increases for higher levels of cognitive-skills. Figure 3 shows kernel densities of non-cognitive skills for deciles 5, 6, 7 and 8 of the cognitive skills' distribution at $t = 1$. Hence, socio-emotional abilities, although positively correlated with cognitive skills, are less so for smarter kids.

6.2 Investment Factors

As explained in Section 4, skills are not the only unobserved characteristics that enter the model. Investment choices made by the family are also unobserved factors. Hence, I estimate the underlying distributions from which the unobserved heterogeneity in investment comes from. That is, I estimate the measurement system described by equations (12) and (13).

The manifest variables used as measures of investment depend on the type of skill they intend to develop. 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 is the child beaten, physically hurt, yelled at or addressed in an inappropriate manner by the parents. Parental control relates to how well parents know where the kid is, who is she with, what is she doing and when is she coming back home. Parental harmony collects information related the level of care and interest in her life the kid feels from her parents.[25] The measures used to identify the cognitive skill investment factor relate to the enrollment in private tutoring of each kid. South Korean society is characterized by the high importance it gives to academic success. Hence, it is not uncommon for kids to be enrolled in after-school academic programs. By age 14, around four fifths of the kids in the survey attend some kind of tutoring. As manifest variables of cognitive skill investment I use a scale of how private —meaning how many classmates there are in every tutoring session, in a reversed scale— the tutoring is, the time spent in tutoring, and the cost of the tutoring.[26]

Tables 5 and 6 present the estimation of the investment in non-cognitive and cognitive skill factors respectively. The results show that the non-cognitive investment factor identified closely relates with good parental practices as it correlates positively with parental control and negatively with physical and verbal abuse. In the same way, the cognitive investment factor identified relates with the quality of after-class tutoring. It is positively correlated with how private the tutoring is, and how many hours the student spends in such after-class activities.

The non-cognitive investment distributions identified, presented in Figure 4, show a remarkable stability of the factor across wave. On the other hand, the cognitive investment distributions, presented in Figure 5, show two important characteristics. First, they are bimodal. That is the case because there are a proportion of kids that take no tutoring at all. Second, they are not stable

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

[26]The information used to inquire about investment in the development of cognitive skills related entirely on the usage of private tutoring. The first score, named type of tutoring, collect information of the nature of the extra-school classes taken. That is, whether the classes were entirely private, with few classmates, with many classmates, or through the internet. Students gave this type of information about their tutoring for every subject (e.g., language, math, science), and based on that I created aggregated measures. The second and third score used were straightforward: the amount of time and money spent in tutoring respectively.

in time. This responds to the fact that participation private tutoring falls as kids grow up.

6.3 Results from the Dynamic Model

6.3.1 Incidence of victimization

Tables 7 and 8 show the relation between skills and selection into bullying.[27] In line with the results of Sarzosa and Urzua (2013a), kids with less non-cognitive skills are significantly more likely to be bullied. Based on these estimates, the model allows me to quantify the probability of selection into bullying for each combination of skills at a given point in time. Figure 6a does this for $t = 1$ (age 14). It 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 have around a 50% chance of being bullied, while those chances for kids in the first decile is virtually zero. An standard deviation increase in non-cognitive skills reduces the likelihood of being bullied by 22 percentage points for the average student. That is, such increase in skills will reduce practically to zero the chances of being victimized.

Table 7 also shows the importance the relation between own and peer characteristics has in determining peer victimization. Controlling for observable characteristics and skill levels, youths who were placed in a school in which their non-cognitive skills are uncommon are significantly more likely to be bullied. The results indicate that the likelihood of victimization of the average student drops by one percentage point for each additional classmate with similar non-cognitive skill endowments he has. Interestingly, the same occurs in terms of income. Bullying probability

[27]Given that the kids are already 14 years old by the first time they are interviewed, there is a possibility for the existence of joint causality between the contemporaneous measures of bullying and skills. I address this issue using the framework described in Hansen et al. (2004) and the exogenous variation that comes from the allocation of students to schools and classrooms. Hansen et al. (2004) require two additional assumptions for identification. First, the assumption of separability between the observed and unobserved part in every equation of the measurement system. Second, the assumption of orthogonality across the error terms in the complete measurement system. This last 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 full measurement system.

falls by half a percentage point for each additional classmate that has a family income level that falls within the epsilon-ball defined around the family income level of the prospective victim.

As I showed in Table 2, the incidence of bullying falls as the kids become older. In fact from wave one to wave two the incidence of being bullied is cut by more than half, falling from 22.5% to 11.12%. However, the results in Table 8 and in Figure 6b show that victimization is not reduced across the board. In fact, while the probability of being bullied has fallen virtually to zero for the kids in the last seven non-cognitive skill deciles, the kids in the first decile remain with similar probabilities of being bullied than a year before, around four times the average. Hence, it is fair to say that in time bullying becomes more selective of its victims, overwhelmingly focusing on kid that lack non-cognitive abilities.

Table 8 indicates that the relation between own and peer characteristics becomes even more important at age 15. Non-cognitive skills uncommonness significantly increases victimization likelihood even more so than before. My results indicate that bullying probability falls by 1.7 percentage points per each additional classmate whose non-cognitive skills fall within the epsilon-ball defined around the non-cognitive skills of the prospective victim.

An important distinction with the empirical strategy in Sarzosa and Urzua (2013a) is that in the present paper I allow for correlated skills. Once I allow for this correlation to exist, I find a positive effect of cognitive skills on the likelihood of being bullied; an effect that is not reported in Sarzosa and Urzua (2013a). This is a very interesting finding as it reflects the fact that the kids that are more likely to be bullied are those who are smart but lack non-cognitive skills. However, Figures 6a and 6b show that any positive effect cognitive skills may have is dwarfed by the negative effect of non-cognitive skills.

6.3.2 Skills Production

Table 9 presents the results of estimating the system described by (1). In the first part of the table, I present the raw estimates of the dynamic parameters. Using these structural parameters —including the one related to selection into bullying, I am able to fully recreate the dynamic process. This process is presented in Figures 7a, 7c, 8a and 8c. Figures 7a and 8a show that high non-cognitive skills produce high future non-cognitive skills, and that marginal increments of those initial skills are very productive (i.e., non-cognitive skills self-productivity $\partial\theta_{t+1}^{NC}/\partial\theta_t^{NC} > 0$ for the entire $\left(\theta_t^{NC}, \theta_t^C\right)$ space, see Figure 9a). These figures also demonstrate that cognitive skills are unimportant in non-cognitive skill production process except for the fact that higher initial cognitive skills make the marginal increments of the initial non-cognitive skills more productive (i.e., $\partial^2\theta_{t+1}^{NC}/\partial\theta_t^{NC}\partial\theta_t^C > 0$).

Figures 7c and 8c show that the production of cognitive skills relies heavily on past levels of cognitive skills. Although the existing levels of non-cognitive skills contribute in the production process of cognitive skills, their contribution is small compared to that of the existing stock of cognitive skills. For instance, going from decile 1 to decile 10 in non-cognitive skills distribution has the same effect on the production of cognitive skills as increasing the cognitive skills input by one decile. Figure 9c shows that, at age 14 (i.e., $t = 1$), the self-productivity of cognitive skills is higher among non-cognitive skilled people. Analogously, Figure 9d shows that the productivity of non-cognitive skills in producing next period cognitive skills is higher among kids with high initial cognitive skills.

My results indicate that there is a strong path dependence in which skills produce skills, setting a high cost in terms of future stock of skills for those who start the accumulation process in the lower quantiles of the skill distribution. My results also show that this path dependence is not reversed by investment choices. In fact, Table 9 and Figure 11 show that investment choices in non-cognitive skills depend greatly on the past level of non-cognitive skills, and investment choices in cognitive skills depend greatly on past levels of that skill in the first place. Hence, people with high skills not only pass their high stock on to the next period, but also they are more prone to invest in their development.

6.3.3 Effects of Bullying on Skill Production and Future Bullying

The right panel of Table 9 shows the effect of bullying on the accumulation of cognitive and non-cognitive skills. To calculate this, I compare the next period skills of those who would be selected into bullying with those who would not, evaluated at the skills' mean. That is, $E\left[\hat{\theta}^{S}_{t+1}|\bar{\theta}^{NC}_{t}, \bar{\theta}^{N}_{t}, M_{t+1}=1\right] - E\left[\hat{\theta}^{S}_{t+1}|\bar{\theta}^{NC}_{t}, \bar{\theta}^{N}_{t}, M_{t+1}=0\right]$ for $S = \{NC, C\}$. For age 14, I find there is a statistically significant effect of -0.0631 of bullying on next period non-cognitive skill accumulation. In order to put into perspective this figure, note that the standard deviation of θ^{NC}_{t+1} is 0.4461. Therefore, my results indicate that bulling reduces non-cognitive skill accumulation by 14.15% of a standard deviation for the average kid. This is a very sizable effect. It implies a reduction of 19.11% of a standard deviation in the language test score, and a reduction of 16.41% of a standard deviation in the math test score. The differential effect of bulling in next period non-cognitive skills depending on previous skills levels is presented in Figure 12a. It shows that this effect can be twice as big for youths with low initial levels of non-cognitive skills.

The same estimation shows there is no statistically significant effect of bullying on cognitive skill accumulation at that age. 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.

Due to the fact that skill levels are important determinants of future bullying, the documented skill depletion caused by bullying is translated into higher probabilities of being victimized again. My dynamic model allows me to see the difference in the incidence of victimization conditional on past victimization. That is, I am able to compare the probability of being bullied at $t = 2$ conditional on having being bullied at $t = 1$. The results presented in Figures 13a and 13b indicate that, despite the overall decrease of bullying incidence, previous victims are more likely to be bullied at $t = 2$ than previous non-victims regardless of their skill levels. The probability of being bullied at $t = 2$ of the previous non-victims is around one fourth less than that of the one that were previously victimized.

The dynamic analysis for age 15 (i.e., $t = 2$) finds that bullying not only becomes more selective, but also more costly. Its effect on the accumulation of non-cognitive skills now reaches -0.3763 for the average student, which represents a reduction of 72.8% of a standard deviation of next period skills. This implies loosing around half of a standard deviation in the language and math test scores. Figure 12b shows that at $t = 2$, bullying is also much more costly to those who lack non-cognitive skills. In particular, the stock of non-cognitive skills lost to bullying by the kids in the first decile of the non-cognitive skill distribution is around 1.5 times greater than the loss of those in the top decile of the distribution.

By age 15, and probably because bullying has became more selective, I do observe statistically significant effect on cognitive skill accumulation. In fact, I estimate that the average kid loses 35% of a standard deviation of next period cognitive skills. This represents loosing 16.64% of a standard deviation in the language test, and 13.42% of a standard deviation in the math test. As in the case of non-cognitive skill accumulation, the size of the effect depends on the skill levels at the beginning of the period. Figure 12d shows that the kids that arrive to this age with low cognitive skills feel the greatest impact of bullying on cognitive skill accumulation. In particular, the cognitive unskilled might lose up to 70% of a standard deviation in next period cognitive skills.

In sum, the average kid will score around two thirds of a standard deviation less in the language and math tests by age 16 if bullied during the last year. These skill loses can also be translated into sizable effect in terms of other outcomes besides test scores.[28] For instance, the average kid would be 28.6 percentage points more likely to report being sick recently, 8.3 percentage points more likely to smoke and 15.8 more likely to drink alcoholic beverages. The stock of skills lost also translates to setbacks in mental health. They equate to an increase of 134.5% of a standard deviation in the depression symptom scale, an increase of a full standard deviation in the levels of stress cased by insecurities regarding his or her image, and 95% of a standard deviation in the levels of stress caused by issues regarding school.

The full path for the average student. My model allows me to track all the process triggered by bullying and in that way measure the gap that widens as time goes by and new victimization events materialize. For instance, let us track the average individual, call it V. Being the average individual, V has an initial set of skills of $\left(\theta_t^{NC}, \theta_t^{C}\right) = (0,0)$. V has a 22.5% chance of being bullied. If V is bullied, V will lose 14.15% of a standard deviation of next period non-cognitive skills. V's cognitive skills remain unchanged. The fact that V was bullied increases V's probability of being bullied again by one fourth compared to the scenario when he was not bullied in the first place. If V is bullied again, let us call this scenario V_{11}, V would have lost 91.65% and 38.47% of a standard deviation of non-cognitive and cognitive skills respectively. If V is not bullied again but was bullied before, scenario V_{10}, V's skill loses would be of 17.23% of a standard deviation of non-cognitive skills and 2.8% of a standard deviation of cognitive skills. If V was not bullied before but is bullied now, scenario V_{01}, V's losses would amount to 72.8% of a standard deviation in the non-cognitive dimension and 35% of a standard deviation in the cognitive one. These results compare to the situation V_{00} when V was never victimized, and therefore V's set of skills remained at $(0,0)$.

Using the results presented in Appendix E to translate these gaps into more understandable

[28] In Appendix E, I run models of unobserved heterogeneity at age 16 on several outcomes like depression, stress in different situations, and the likelihood of smoking, drinking alcohol, felling healthy, being satisfied with life and going to college by age 19.

metrics, we see that for V_{11} they represent a 36 percentage points increase in the probability of feeling sick, a 10.3 percentage points increase in the likelihood of smoking, 19.8 percentage points increase in the likelihood of drinking alcohol. They also represent an increase of 170% of a standard deviation in the depression symptom scale, an increase of 133% of a standard deviation in the stress due to image scale, and a 120% of a standard deviation increase in the stress caused by school.

Note that this exercise follows the average student. Therefore, around half of students in the sample will face harsher consequences than the ones just described due to the fact that they will start this middle-school journey with lower stocks of skills.

6.3.4 Complementarieties

As explained in Section 3, the analysis of the static and dynamic complementarieties between skills and bullying allows the measurement of how much the effect of the shock on skill formation is modified by a marginal change in previous period skills, and to what extent the effect of bullying is compounded on past bullying events. My results regarding the static complementariety on bullying (i.e., equation (2)), presented in Figures 14 and 15, show that marginal increases of the level of non-cognitive skills drastically reduce the negative effect of bullying on non-cognitive skill accumulation. In fact, comparing Figures 12a (i.e., the size of the effect on bullying at age 14 on non-cognitive skill formation at age 15) and 14a we see that a marginal increase in non-cognitive skills would have reduced the negative effect of bullying by around one half. For the average kid, the raw effect of bullying on non-cognitive skills at age 15 would be reduced from -0.0631, to -0.0368 by just a marginal increase in the previous period non-cognitive skills. At age 16, the static complementariety between bullying and non-cognitive skills represents around one third of the full effect of bullying on non-cognitive skill formation. Therefore, marginal increases in the previous period non-cognitive skills would have brought down the size of the bullying effect for the average kid from -0.3763 to -0.2414. These sizable reductions contrast with the evidence presented in

Figures 14b and 15b that demonstrate the negligible influence that marginal increases of cognitive skills have on reducing the impact of bullying on non-cognitive skill accumulation.

Regarding the effect bullying has at age 15 on cognitive skill accumulation, Figures 15c and 15d indicate that marginal increases in the stock of both cognitive and non-cognitive skills lessen its negative effect.[29] For the average kid, the effect of bullying on cognitive skill production would fall by a third due to marginal increases in previous period non-cognitive skills, and two thirds due to a marginal increase in the previous period stock of cognitive skills.

The analysis on the dynamic complementariety of bullying on skill accumulation (i.e., equation (3)), presented in Figure 16, shows that there is, in fact, a compounded effect of bullying, especially for those with low non-cognitive skills. That is, the negative effect of bullying at age 15 on skill accumulation at age 16 is greater if the person was bullied at age 14. Hence, not only bullying becomes more selective on low non-cognitive skilled people, as shown in Figures 6a and 6b, but also a second bullying event itself is more harmful.

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

[29]Figures 14c and 14d present the static complementariety analysis for cognitive skill accumulation at age 14. They are presented for the sake of completeness. I will not thoroughly analyze their results because there was no effect of bullying on that skill dimension for that age in the first place.

7 Policy Implications

Several anti-bullying campaigns have been deployed all around the world in an effort ambitious effort to eliminate this unwanted phenomenon. Prominent psychologists and governmental institutions are continuously involved in the development of programs to deter bullying.[30] My findings indicate there are at least two fronts policymakers can work on. First, the development of non-cognitive skills. Non-cognitive skilled kids will not only be more likely to be successful adults (Duckworth and Seligman, 2005; Heckman et al., 2006b), but also they are dramatically less likely to be victimized. And if —despite the low probability of being so— they happen to be bullied, its impact on their skill accumulation path is much lessened. The importance of developing non-cognitive skills at a young age is heightened by the strong dependance of current skill levels on past skills levels.

The second implication of my results relates with classroom assignment. Tables 7 and 8 show that, given the skill levels, children with uncommon characteristics are more likely to be targeted by bullies. Therefore, in an effort to illustrate how much would bullying be reduced if kids were less likely to be found in classroom where some of their characteristics end up being uncommon, I simulate a different type of classroom allocation. One that is unfeasible in practice and treated as a benchmark, places all the kids in the survey in classroom with kids that have similar stocks of non-cognitive skills. This exercise ignores geographical distances. It just sorts the universe of students with respect to their non-cognitive skills and split them in classrooms according to the typical classroom size in South Korea.

The results of these simulations are presented in Figure 17. As in Figure 6a, it plots the likelihood of being bullied for every skill level. A comparison between these two figures shows the massive impact that reducing in-classroom non-cognitive skill heterogeneity has on the likelihood of being victimized. The benchmark case presented in Figure 17 shows that by arranging students

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

with classmates that have similar levels of non-cognitive skills, the overall likelihood of victimization falls from 22.5% to 5.5%. This dramatic reduction is focused on medium and high non-cognitive skilled students, for which the hazard of being bullied almost disappears. Although proportionally less, there is also a substantial reduction of victimization among low non-cognitive students. If placed in skill homogenous classrooms, low non-cognitive skilled students would see their chances of being victimized drop by half.

8 Conclusions

This paper develops and estimates a structural model of skill accumulation that explicitly introduces a peer-affected input. This way, I introduce bullying as a skill depleter event into a model that is general enough to allow past skills, parental investment choices and bullying to affect future stock of skills, and at the same time investment choices and bullying to be affected by past skill levels. The model uses several dimensions of unobserved heterogeneity and in-classroom variation of student characteristics to identify the victims. My findings indicate the existence of a vicious cycle between victimization and skill depletion. I find that bullying is disproportionately suffered by students that lack socio-emotional skills, and among those, the smart students are more likely to be victimized. My findings, in line with psychological studies, suggest that conditional on the level of skills, kids with uncommon characteristics relative to those of their classmates are more likely to be victimized.

The dynamic estimation showed that bullying is very costly in terms of the amount of skills lost from one period to the next. Bulling at age 14 reduces non-cognitive skill accumulation by a 14.15% of a standard deviation for the average kid. That effect is twice as big for kids with low initial levels of non-cognitive skills. At age 15, bullying not only becomes more selective, but also more costly. It reduces next period skills by almost three-fourths of a standard deviation. Static complementariety shows current stock of non-cognitive skills—unlike the cognitive one—influences greatly the "negative productivity" of the bullying event. In the same vein, the analysis of dynamic complementariety shows that the effect of being bullied at age 15 on the skills at age 16 is greater for those who were also bullied at age 14, especially those who started with low non-cognitive skills.

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

This paper intends to contribute to the human development literature in economics with explanations of how victimization of school-aged kids may hamper the development of successful adults. In this process, this paper contributed to the skill formation literature by introducing latent factor dependent shocks as triggers of phenomena that have long-lasting consequences. In this context, this paper opens a promising research agenda that can continue in at least two ways. First, by extending my model to incorporate additional characteristics and dynamics regarding bullying, and second, by analyzing the skill accumulation consequences of other types of shocks. Among the latter, new research is needed to analyze the consequences of shocks like parental separation or loss, and health or family financial mishaps. Among the former, promising research opportunities arise in the analysis of the role that gender plays in classroom dynamics *vis-a-vis* bullying, or—data permitting—the introduction of physical traits as determinants of victimization. Furthermore, given the importance of initial levels of skills, it is crucial to explore how bullying affects skill accumulation among younger cohorts.

References

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

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

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

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

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

Boulton, M. and Underwood, K. (1993). Bully/victim problems among middle school children. *European education*, 25(3):18–37.

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.

Bukowski, W. M., Hoza, B., and Boivin, M. (1995). Friendship protects "at risk" children from victimization by peers. *The role of friendship in children's developmental risk and resilience: A developmental pathology perspective*.

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.

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

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *IZA Discussion Paper Series*, 2550.

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

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

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

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

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

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

Eriksen, T., Nielsen, H., and Simonsen, M. (2012). The effects of bullying in elementary school. *Economics Working Papers*, 16.

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*.

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

Heckman, J., Urzua, S., and Vytlacil, E. (2006b). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics*.

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.

Hodges, E. and Perry, D. (1996). Victims of peer abuse: An overview. *Relaiming Chldren and youth*, Spring:23–28.

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*.

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

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. (2005). School bullying and suicidal risk in korean middle school students. *PEDIATRICS*, 115(2):357–363.

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.

Kochenderfer, B. and Ladd, G. (1996). Peer victimization: Cause or consequence of school maladjustment? *Child development*, 67(4):1305–1317.

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

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

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

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

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

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

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

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

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

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

of Education, U. D. (2011). Analysis of state bullying laws and policies. pages 1–202.

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

Ouellet-Morin, I., Odgers, C. L., Danese, A., Bowes, L., Shakoor, S., Papadopoulos, A. S., Caspi, A., Moffitt, T. E., and Arseneault, L. (2011). Blunted cortisol responses to stress signal social and behavioral problems among maltreated/bullied 12-year-old children. *BPS*, 70(11):1016–1023.

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)*.

Perry, D., Kusel, S., and Perry, L. (1988). Victims of peer aggression. *Developmental psychology*, 24(6):807.

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

Prada, M. F. and Urzua, S. (2013). One size does not fit all: The role of vocational ability on college attendance and labor market outcomes. *MIMEO Univerity of Maryland*.

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.

Sarzosa, M. and Urzua, S. (2012). Implemeting factor models in stata: the heterofactor command. *MIMEO. University of Maryland*, pages 1–6.

Sarzosa, M. and Urzua, S. (2013a). Bullying and cyberbullying in teenagers, the role of cognitive and non-cognitive skills. *University of Maryland*, pages 1–56.

Sarzosa, M. and Urzua, S. (2013b). Cognitive and non-cognitive skills in teenagers: Happinness, cybercrime, bullying and their dynamics in south korea. *University of Maryland*.

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

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

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

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

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

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

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

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

Tough, P. (2012). How children succeed: Grit, curiosity, and the hidden power of character.

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

Vossekuil, B., Fein, R., Reddy, M., Borum, R., and Modzeleski, W. (2002). The final report and findings of the safe school initiative. *US Secret Service and Department of Education, Washington, DC*.

Table 1: Descriptive Statistics

Total sample size	3,449		
Number of Females	1,724	Fathers Education:	
Proportion of urban households	78.55%	High-school	42.94%
Prop. of single-headed households	6%	4yr Coll. or above	36.56%
Median monthly income per-capita	1mill won	Mothers Education:	
Prop. of Youths in College by 19	56.65%	High-school	56.31%
Incidence of smoking by 19*	19.08%	4yr Coll. or above	19.51%
Prop. of Single-child households	8.6%		

*Incidence calculated as the proportion of people who has smoked at least once in the last year

Table 2: Incidence of Bullying by Wave

	Bullied	Bullied in $t=1$	
Wave		No	Yes
1	.22499	·	·
2	.11198	.07377	.24271
3	.04768	.03270	.16871
4	.03428	.02275	.25517
5	.02231	.01941	.11340

Table 3: Identification of Skills at $t = 1$

VARIABLES	(1) Locus	(2) Irrespons	(3) Self-est	(4) Lang-SSc	(5) Math-Scie	(6) YearExam
Age (months)	-0.009*	0.015***	-0.014***	-0.008*	-0.009**	-0.014***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Male	0.150***	-0.055	0.170***	0.024	0.312***	-0.037
	(0.036)	(0.036)	(0.036)	(0.031)	(0.032)	(0.026)
Older Siblings	0.017	-0.006	-0.017	-0.045	0.032	0.009
	(0.033)	(0.033)	(0.034)	(0.029)	(0.030)	(0.026)
Young Siblings	0.016	-0.069**	0.023	0.077**	0.086***	0.086***
	(0.035)	(0.034)	(0.035)	(0.030)	(0.031)	(0.026)
lnInc_pc	0.074**	-0.104***	0.022	0.150***	0.145***	0.126***
	(0.033)	(0.033)	(0.033)	(0.029)	(0.029)	(0.025)
Urban	0.172***	-0.084	0.065	0.101**	0.067	-0.016
	(0.053)	(0.053)	(0.054)	(0.045)	(0.046)	(0.036)
Lives: Both Parents	0.209**	-0.313***	0.275***	0.314***	0.369***	0.235***
	(0.102)	(0.101)	(0.103)	(0.091)	(0.093)	(0.086)
Lives: Only Mother	0.325**	-0.274**	0.428***	0.282**	0.354***	0.072
	(0.136)	(0.135)	(0.137)	(0.120)	(0.123)	(0.110)
Father Edu: 2yColl	0.133*	-0.147**	-0.022	0.145**	0.197***	0.205***
	(0.072)	(0.071)	(0.072)	(0.061)	(0.063)	(0.051)
Father Edu: 4yColl	0.141***	-0.142***	0.087**	0.317***	0.187***	0.245***
	(0.043)	(0.042)	(0.043)	(0.037)	(0.038)	(0.031)
Father Edu: GS	0.263***	-0.321***	0.148*	0.464***	0.288***	0.317***
	(0.076)	(0.075)	(0.076)	(0.063)	(0.065)	(0.049)
Non-Con. Factor	1.213***	-1.258***	1	0.864***	0.934***	
	(0.090)	(0.096)	.	(0.110)	(0.115)	
Cognitive Factor				0.533***	0.503***	1
				(0.026)	(0.027)	.
Constant	-0.755***	0.837***	-0.408**	-1.128***	-1.273***	-0.786***
	(0.174)	(0.173)	(0.175)	(0.157)	(0.160)	(0.148)
Observations	3,097					

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals

Table 4: Identification of Skills at 2

VARIABLES	(1) Locus	(2) Irrespons	(3) Self-est	(4) Lang-SSc	(5) Math-Scie	(6) YearExam
Age (months)	-0.023***	0.008	-0.008	-0.013***	-0.008*	-0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Male	0.112***	-0.067*	0.136***	0.059*	0.366***	-0.058**
	(0.038)	(0.038)	(0.038)	(0.033)	(0.033)	(0.026)
Older Siblings	0.034	-0.016	0.055	-0.004	0.013	0.023
	(0.036)	(0.036)	(0.036)	(0.032)	(0.032)	(0.027)
Young Siblings	0.041	-0.092**	0.078**	0.143***	0.125***	0.082***
	(0.037)	(0.037)	(0.037)	(0.033)	(0.033)	(0.029)
lnInc_pc	0.087**	-0.039	0.068*	0.158***	0.164***	0.171***
	(0.037)	(0.037)	(0.037)	(0.034)	(0.033)	(0.030)
Urban	0.099*	-0.005	0.023	0.084*	0.059	-0.099**
	(0.058)	(0.058)	(0.058)	(0.050)	(0.050)	(0.039)
Lives: Both Parents	-0.079	-0.186**	0.062	0.286***	0.411***	0.302***
	(0.087)	(0.087)	(0.088)	(0.076)	(0.075)	(0.062)
Lives: Only Mother	0.022	-0.244*	0.141	0.068	0.214*	0.220**
	(0.132)	(0.132)	(0.132)	(0.114)	(0.113)	(0.092)
Father Edu: 2yColl	-0.004	-0.209***	0.102	0.088	0.187***	0.178***
	(0.075)	(0.075)	(0.075)	(0.066)	(0.065)	(0.052)
Father Edu: 4yColl	0.112**	-0.166***	0.105**	0.295***	0.219***	0.253***
	(0.045)	(0.045)	(0.045)	(0.039)	(0.039)	(0.030)
Father Edu: GS	0.211**	-0.245***	0.119	0.358***	0.304***	0.346***
	(0.086)	(0.086)	(0.086)	(0.076)	(0.075)	(0.058)
Non-Con. Factor	1.190***	-1.325***	1	1.351***	1.160***	
	(0.109)	(0.131)	.	(0.215)	(0.179)	
Cognitive Factor				0.405***	0.461***	1
				(0.039)	(0.034)	.
Constant	-0.328	0.445**	-0.499**	-1.125***	-1.419***	-0.937***
	(0.200)	(0.201)	(0.201)	(0.181)	(0.180)	(0.163)
Observations	2,731					

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals

Table 5: Identification of Unobserved Non-Cognitive Investment Factor

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

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals

Table 6: Identification of Unobserved Cognitive Investment Factor

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

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals

Table 7: Likelihood of Being Bullied at $t = 1$

VARIABLES	(1)		(2)		(3) Coefficients		(4)		(5)	
Age (months)	0.0154*	(0.009)	0.0134	(0.008)	0.0152*	(0.009)	0.0149*	(0.009)	0.0148*	(0.009)
Male	0.3084***	(0.065)	0.2900***	(0.062)	0.3109***	(0.065)	0.3025***	(0.065)	0.3048***	(0.065)
Older Siblings	-0.0421	(0.058)	-0.0300	(0.055)	-0.0433	(0.058)	-0.0485	(0.058)	-0.0485	(0.058)
Young Siblings	-0.0878	(0.061)	-0.0721	(0.057)	-0.0888	(0.061)	-0.0922	(0.061)	-0.0926	(0.061)
lnInc_pc	-0.1005*	(0.058)	-0.0582	(0.054)	-0.1055*	(0.058)	-0.1221**	(0.059)	-0.1256**	(0.059)
Urban	0.0188	(0.091)	0.0187	(0.088)	0.0321	(0.091)	0.0208	(0.091)	0.0312	(0.091)
Lives: Both Parents	-0.1960	(0.172)	-0.2848*	(0.158)	-0.1863	(0.172)	-0.1819	(0.172)	-0.1729	(0.172)
Lives: Only Mother	-0.1302	(0.228)	-0.1921	(0.212)	-0.1288	(0.228)	-0.1489	(0.228)	-0.1456	(0.229)
Father Edu: 2yColl	0.0411	(0.121)	0.0461	(0.117)	0.0398	(0.122)	0.0434	(0.121)	0.0430	(0.121)
Father Edu: 4yColl	-0.0577	(0.076)	0.0064	(0.072)	-0.0622	(0.076)	-0.0585	(0.076)	-0.0641	(0.076)
Father Edu: GS	0.1803	(0.134)	0.1666	(0.125)	0.1677	(0.134)	0.1673	(0.134)	0.1559	(0.134)
$E_{-i\in c}$[SchoolQty]	-0.0768	(0.112)			-0.0663	(0.112)	-0.0727	(0.111)	-0.0625	(0.111)
$E_{-i\in c}$[Non-Cog]			-0.2018	(0.345)						
$E_{-i\in c}$[Cognitiv]			0.5116	(0.495)						
Mass[Non-Cog]					-0.0401**	(0.019)			-0.0355*	(0.019)
Mass[Cognitive]					0.0257	(0.022)			0.0311	(0.022)
Mass[Income]							-0.0201*	(0.010)	-0.0194*	(0.011)
Non-Cogn	-1.9174***	(0.353)	-1.9029***	(0.333)	-1.9166***	(0.353)	-1.8980***	(0.350)	-1.8995***	(0.352)
Cognitive	0.2870***	(0.070)	0.2881***	(0.066)	0.2770***	(0.070)	0.2840***	(0.069)	0.2726***	(0.069)
Constant	-0.5133*	(0.298)	-0.6007**	(0.281)	-0.4646	(0.309)	-0.3196	(0.313)	-0.3077	(0.320)
					Marginal Effects at the Mean					
Mass[Non-Cog]					0.0115				0.0101	
Mass[Cognitive]					0.0064				0.0076	
Mass[Income]							-0.0057		-0.0055	
Non-Cogn	-0.5061		-0.5125		-0.5061		-0.5009		-0.5005	
Cognitive	0.0757		0.0776		0.0731		0.0749		0.0718	
Observations	2,805		3,097		2,805		2,805		2,805	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Estimations include school fixed-effects. Mass[] refers to the number of observations within a window of 10% of a SD around observation i. The marginal effect of the Mass[] variables are calculated based on the discrete change in the number of people inside the window from 0 to 1.

Table 8: Likelihood of Being Bullied at $t = 2$

VARIABLES	(1)		(2)		(3)		(4)		(5)	
	Coefficients									
Age (months)	0.0077	(0.011)	0.0072	(0.011)	0.0069	(0.011)	0.0071	(0.011)	0.0068	(0.011)
Male	0.3381***	(0.083)	0.2878***	(0.082)	0.3440***	(0.084)	0.3340***	(0.083)	0.3442***	(0.084)
Older Siblings	-0.0233	(0.075)	-0.0643	(0.073)	-0.0174	(0.075)	-0.0217	(0.074)	-0.0176	(0.075)
Young Siblings	-0.1520*	(0.081)	-0.1851**	(0.081)	-0.1437*	(0.081)	-0.1504*	(0.080)	-0.1440*	(0.081)
lnInc_pc	0.0249	(0.078)	-0.0090	(0.076)	0.0182	(0.078)	0.0074	(0.078)	0.0184	(0.078)
Urban	0.0101	(0.120)	0.0115	(0.119)	0.0301	(0.120)	0.0172	(0.119)	0.0307	(0.121)
Lives: Both Parents	0.0305	(0.188)	0.0694	(0.184)	0.0266	(0.189)	0.0509	(0.188)	0.0277	(0.189)
Lives: Only Mother	0.3882	(0.257)	0.3485	(0.255)	0.3941	(0.259)	0.3764	(0.256)	0.3939	(0.259)
Father Edu: 2yColl	0.0347	(0.151)	-0.0006	(0.152)	0.0292	(0.152)	0.0273	(0.151)	0.0291	(0.152)
Father Edu: 4yColl	-0.1319	(0.098)	-0.1199	(0.096)	-0.1419	(0.099)	-0.1335	(0.098)	-0.1417	(0.099)
Father Edu: GS	0.1202	(0.172)	0.1072	(0.169)	0.1116	(0.173)	0.1148	(0.171)	0.1116	(0.173)
$E_{-i\in c}$[SchoolQty]	0.2132	(0.184)			0.2327	(0.184)	0.2168	(0.183)	0.2370	(0.184)
$E_{-i\in c}$[Non-Cog]			-0.4134	(0.548)						
$E_{-i\in c}$[Cognitiv]			0.9342*	(0.536)						
Mass[Non-Cog]					-0.1118**	(0.056)			-0.1118**	(0.056)
Mass[Cognitive]					-0.0397	(0.056)			-0.0392	(0.057)
Mass[Income]							-0.0195	(0.014)	-0.0014	(0.015)
Non-Cogn	-2.3146***	(0.588)	-2.4713***	(0.614)	-2.3480***	(0.592)	-2.2924***	(0.586)	-2.3471***	(0.592)
Cognitive	0.3436***	(0.093)	0.3488***	(0.094)	0.3612***	(0.094)	0.3402***	(0.093)	0.3608***	(0.095)
Constant	-1.7251***	(0.425)	-1.5648***	(0.415)	-1.4890***	(0.437)	-1.5730***	(0.433)	-1.4862***	(0.438)
	Marginal Effects at the Mean									
Mass[Non-Cog]					-0.0178				-0.0177	
Mass[Cognitive]					-0.0056				-0.0055	
Mass[Income]							-0.0030		-0.0002	
Non-Cogn	-0.2922		-0.3269		-0.3152		-0.3176		-0.3146	
Cognitive	0.0463		0.0460		0.0485		0.0471		0.0484	
Observations	2,553		2,725		2,550		2,553		2,550	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Estimations include school fixed-effects. Mass[] refers to the number of observations within a window of 10% of a SD around observation i. The marginal effect of the Mass[] variables are calculated based on the discrete change in the number of people inside the window from 0 to 1.

Table 9: Dynamic Estimation $\theta^S_{t+1} = f\left(\theta^{NC}_t, \theta^C_t, I^S, Bullied\right)$

	$(t=1)$		$(t=2)$	
	θ^{NC}_{t+1}	θ^{C}_{t+1}	θ^{NC}_{t+1}	θ^{C}_{t+1}
θ^{NC}_t	1.0513***	0.3505***	1.4077***	0.3802***
	(0.039)	(0.095)	(0.067)	(0.020)
θ^{C}_t	-0.0569***	0.7367***	-0.1329***	0.8239***
	(0.015)	(0.026)	(0.019)	(0.007)
I^{NC}_{t+1}	0.0570**	0.0525***	0.0447	-0.0166***
	(0.024)	(0.009)	(0.029)	(0.003)
M_{t+1}	-0.0514**	-0.1298*	-0.3195***	-0.1873***
	(0.025)	(0.743)	(0.042)	(0.015)
ρ	0.3755	-0.1431	0.2350*	0.6774***
	(0.385)	(0.1217)	(0.127)	(0.057)
$\alpha^{NC}_{NC,t}$	0.7917***	0.1712	1.1452***	0.5654***
	(0.122)	(0.190)	(0.158)	(0.1955)
$\alpha^{NC}_{C,t}$	-0.033	0.2519***	-0.0421	0.3578***
	(0.031)	(0.046)	(0.036)	(0.052)
Obs.	2,345		2,233	

$E\left[\theta^{NC}_{t+1} \mid \bar{\theta}^{NC}_t, \bar{\theta}^{N}_t, M_{t+1}=1\right] - E\left[\theta^{NC}_{t+1} \mid \bar{\theta}^{NC}_t, \bar{\theta}^{N}_t, M_{t+1}=0\right]$	$(t=1)$	$(t=2)$
θ^{NC}_{t+1}		
Raw Effect	-0.0631***	-0.3763***
	(0.029)	(0.065)
As StdDev of θ^{NC}_{t+1}	-0.1415	-0.7280
θ^{C}_{t+1}		
Raw Effect	-0.1199	-0.2957***
	(0.163)	(0.022)
As StdDev of θ^{C}_{t+1}	-0.1566	-0.3504
Obs.	2,345	2,233

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

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

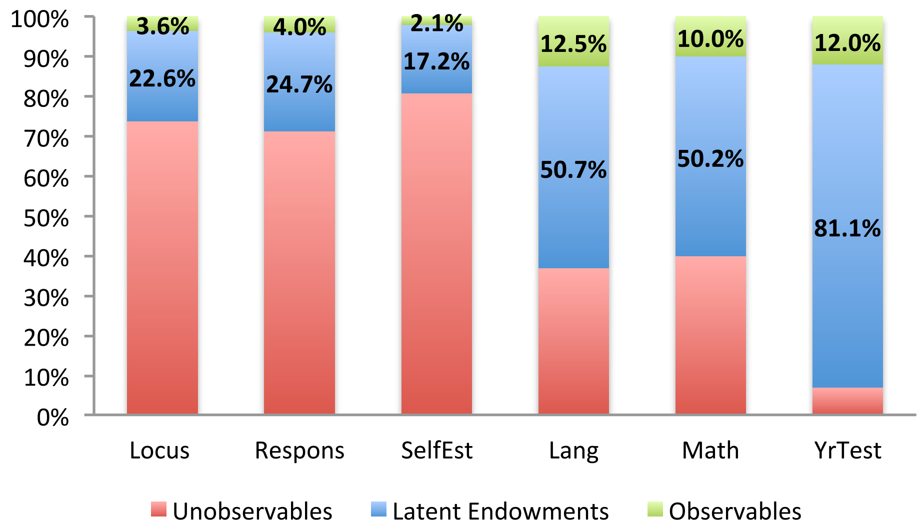
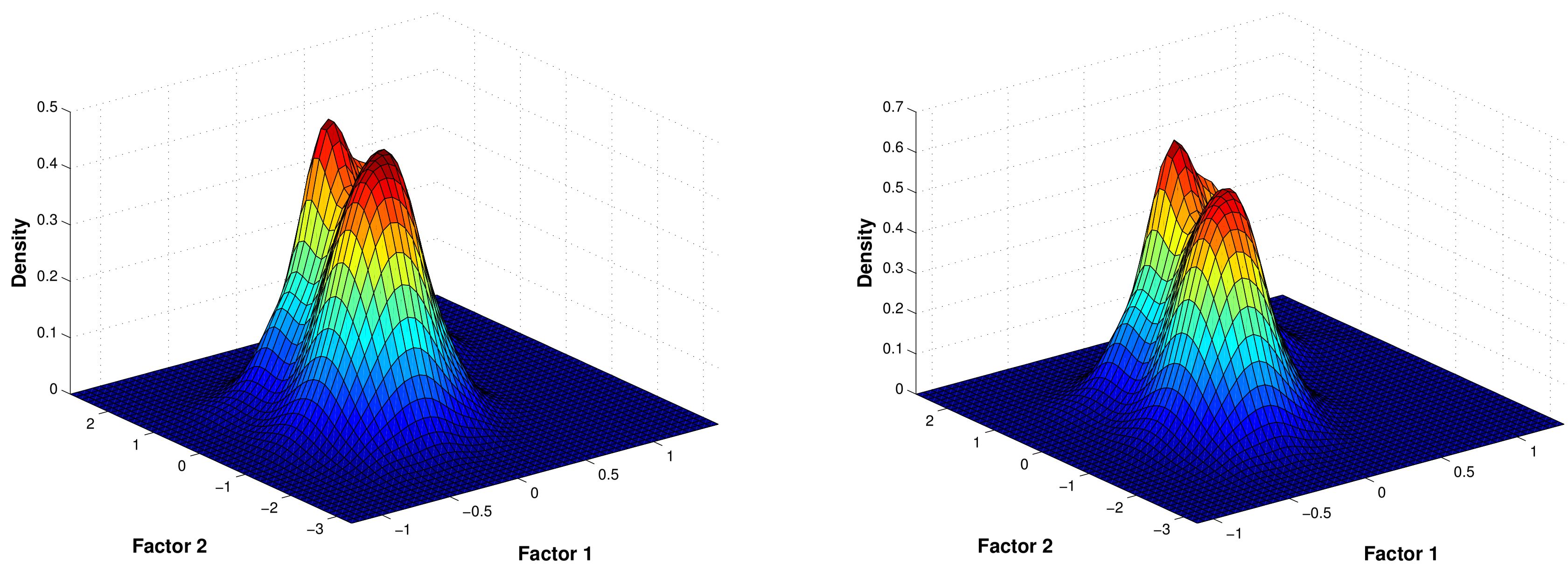


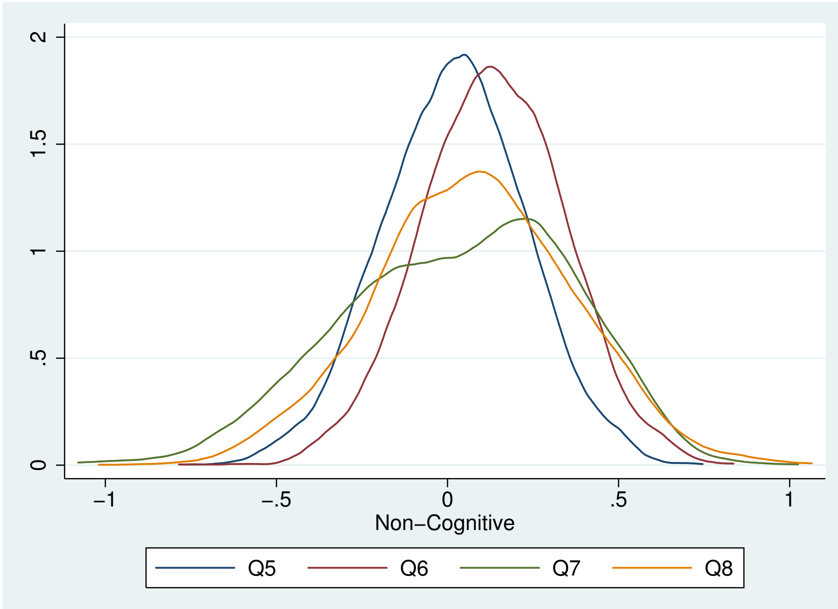
Figure 2: Joint Skills Distribution

(a) $t = 1$ (b) $t = 2$



Note: Factor 1 refers to non-cognitive skills and Factor 2 refers to cognitive skills.

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



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

Figure 4: Unobserved Non-Cognitive Investment Factor

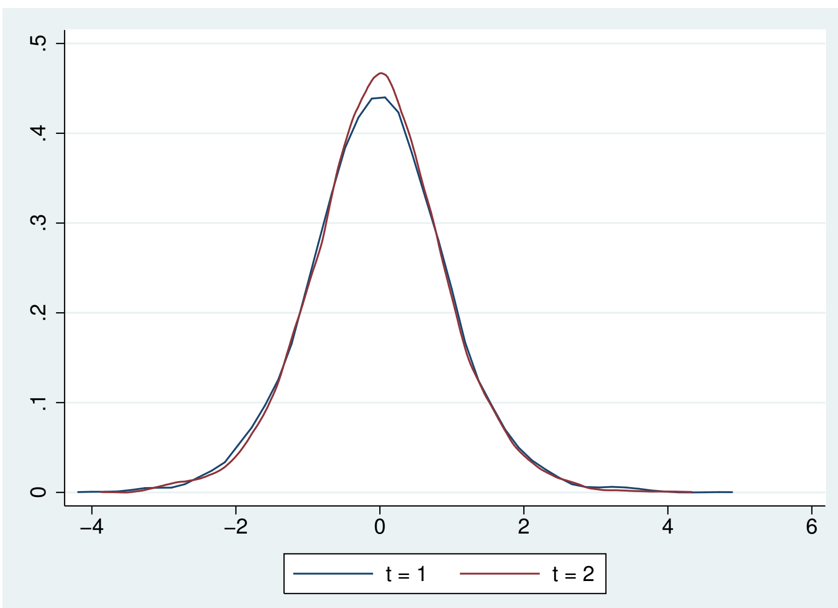


Figure 5: Unobserved Cognitive Investment Factor

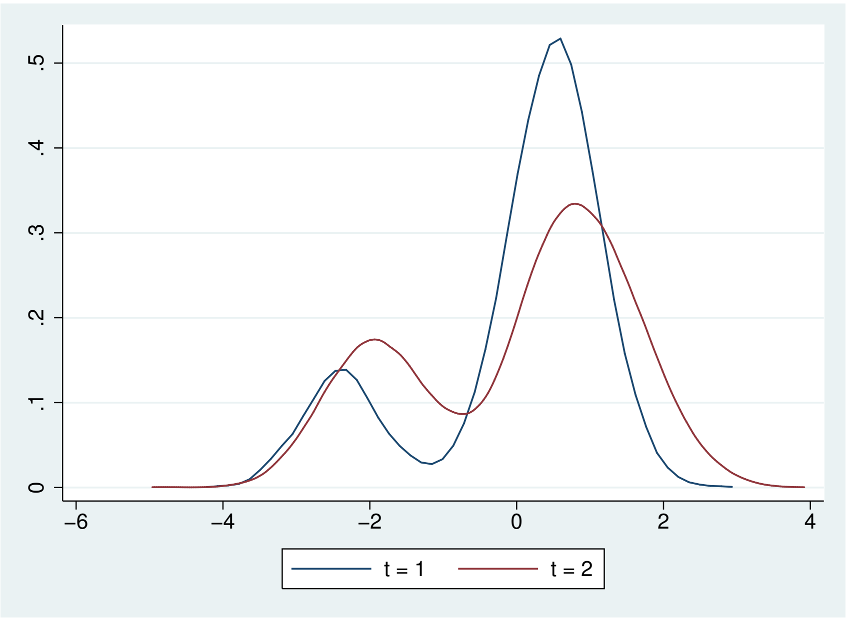


Figure 6: Probability of Being Bullied

(a) $t = 1$ (b) $t = 2$

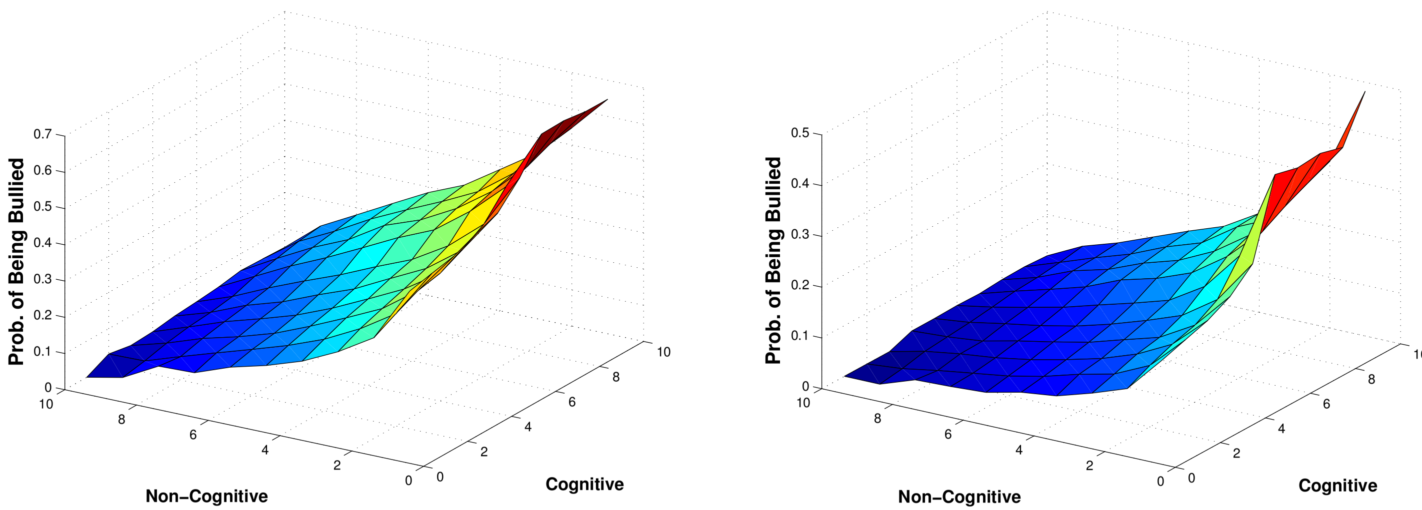
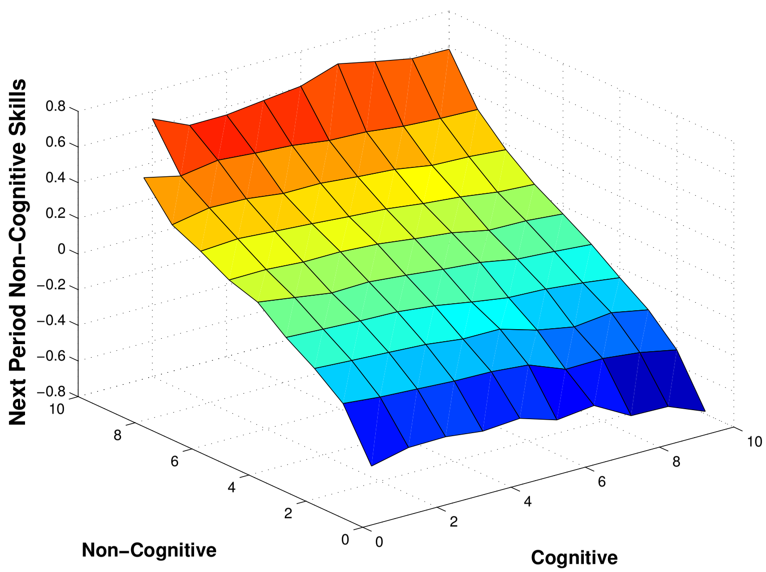
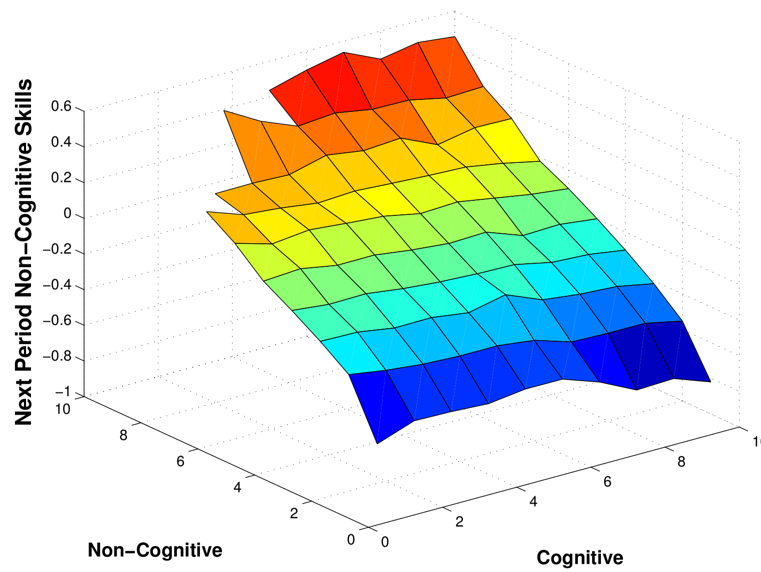


Figure 7: θ_{t+1}^{S} as a function of θ_t^{NC} and θ_t^{C} $(t = 1)$

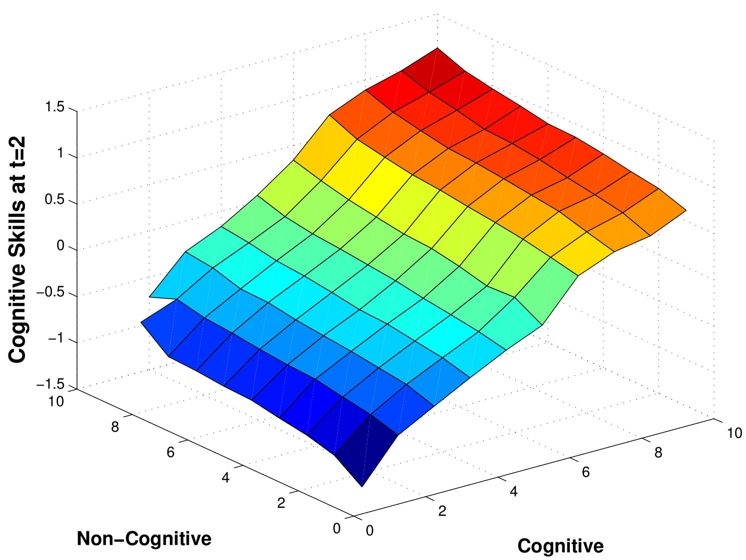
(a) Non-Cognitive: Everyone



(b) Non-Cognitive: Bullying Victims



(c) Cognitive: Everyone



(d) Cognitive: Bullying Victims

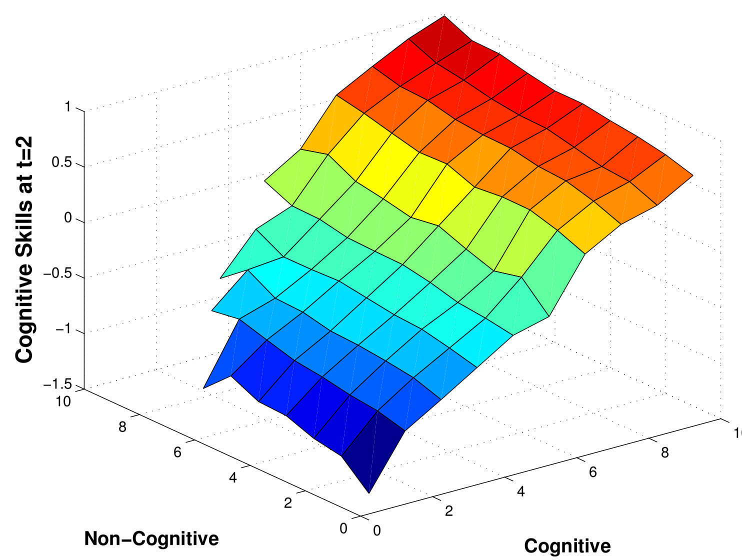
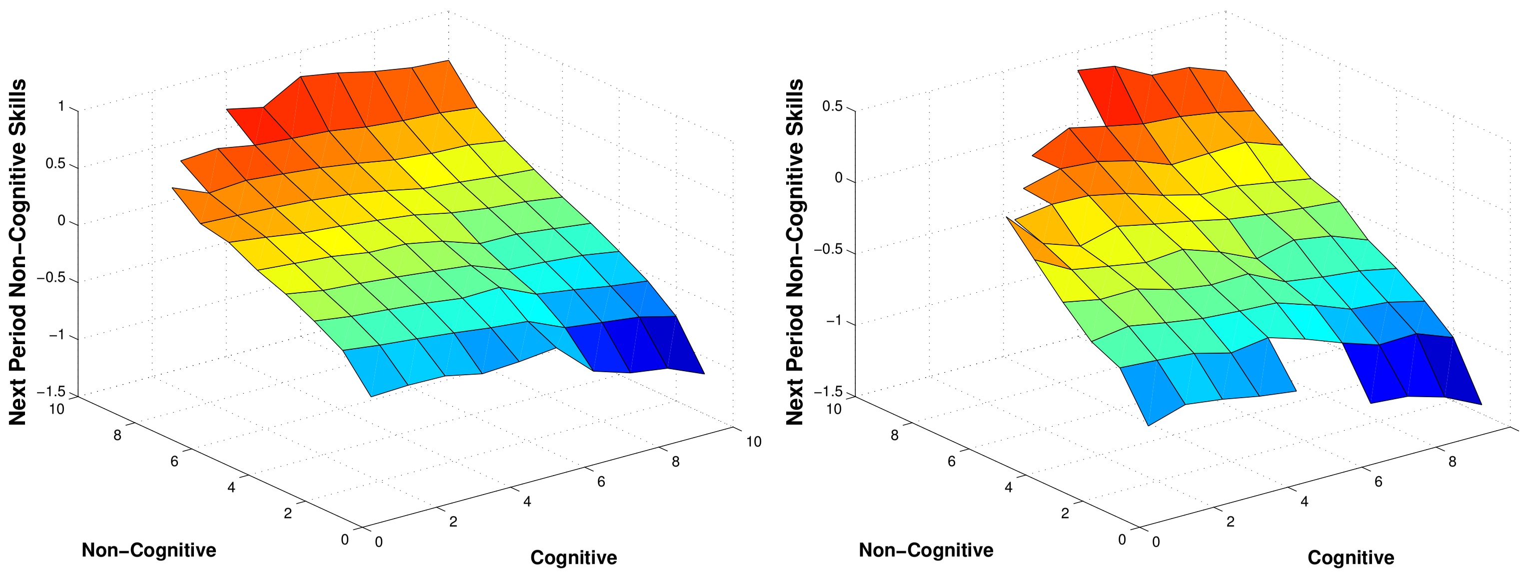


Figure 8: θ^S_{t+1} as a function of θ^{NC}_t and θ^C_t ($t = 2$)

(a) Non-Cognitive: Everyone

(b) Non-Cognitive: Bullying Victims



(c) Cognitive: Everyone

(d) Cognitive: Bullying Victims

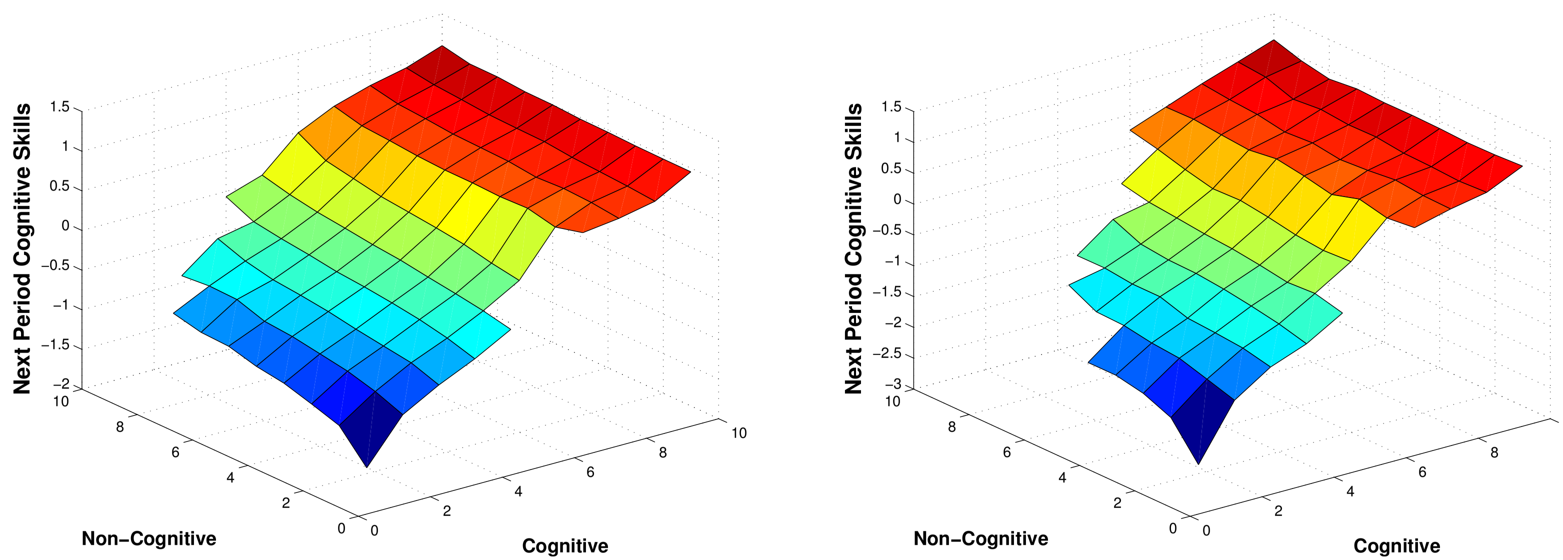
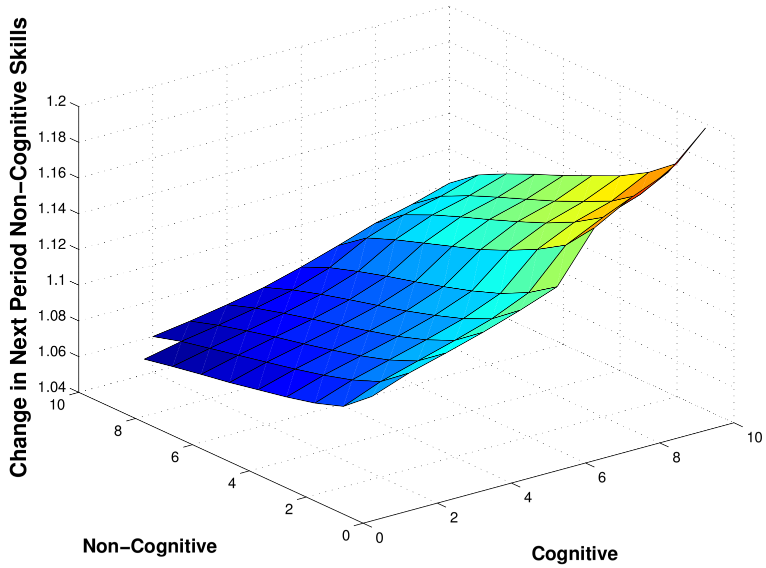
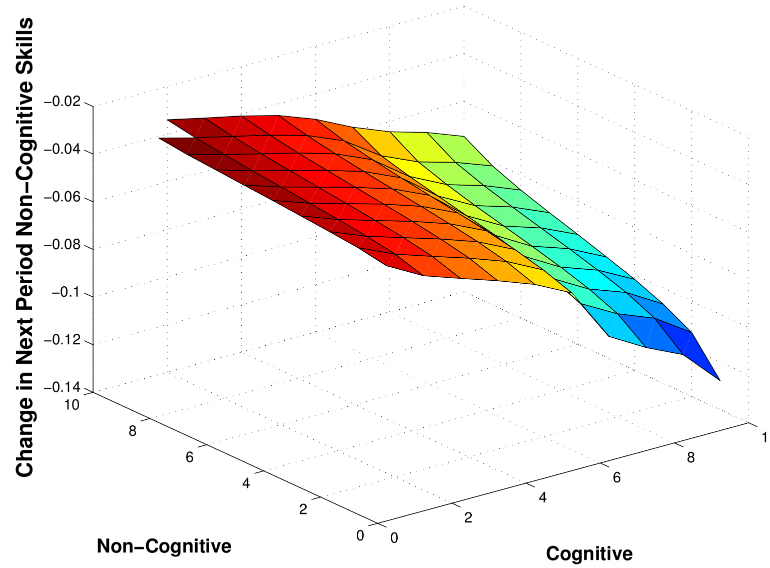


Figure 9: Self ($\partial\theta_{t+1}^{S}/\partial\theta_{t}^{S}$) and Cross ($\partial\theta_{t+1}^{S}/\partial\theta_{t}^{S'}$) Productivities ($t = 1$)

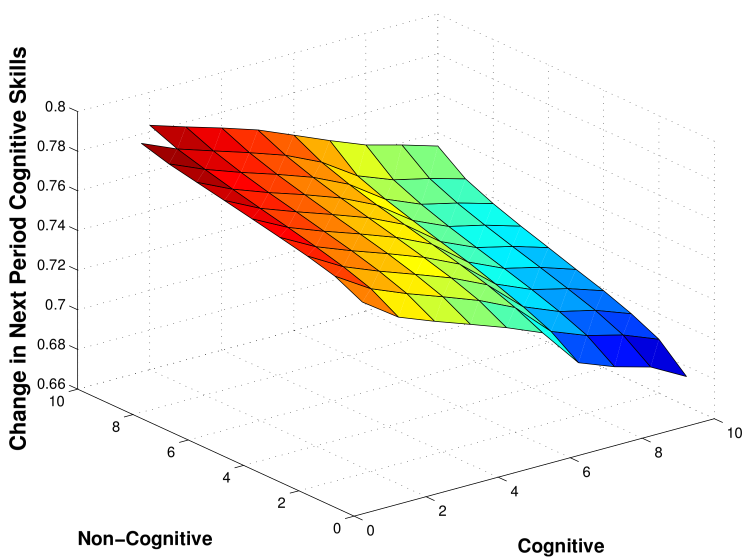
(a) Non-Cognitive: Self Productivity



(b) Non-Cognitive: Cross Productivity



(c) Cognitive: Self Productivity



(d) Cognitive: Cross Productivity

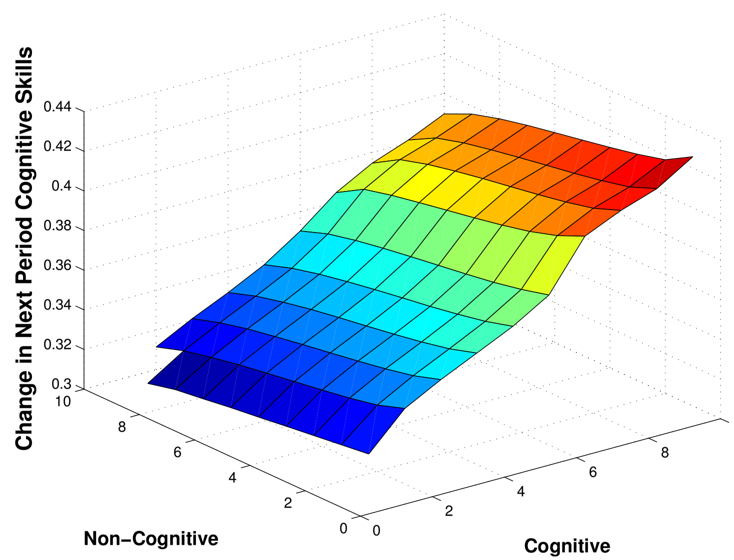
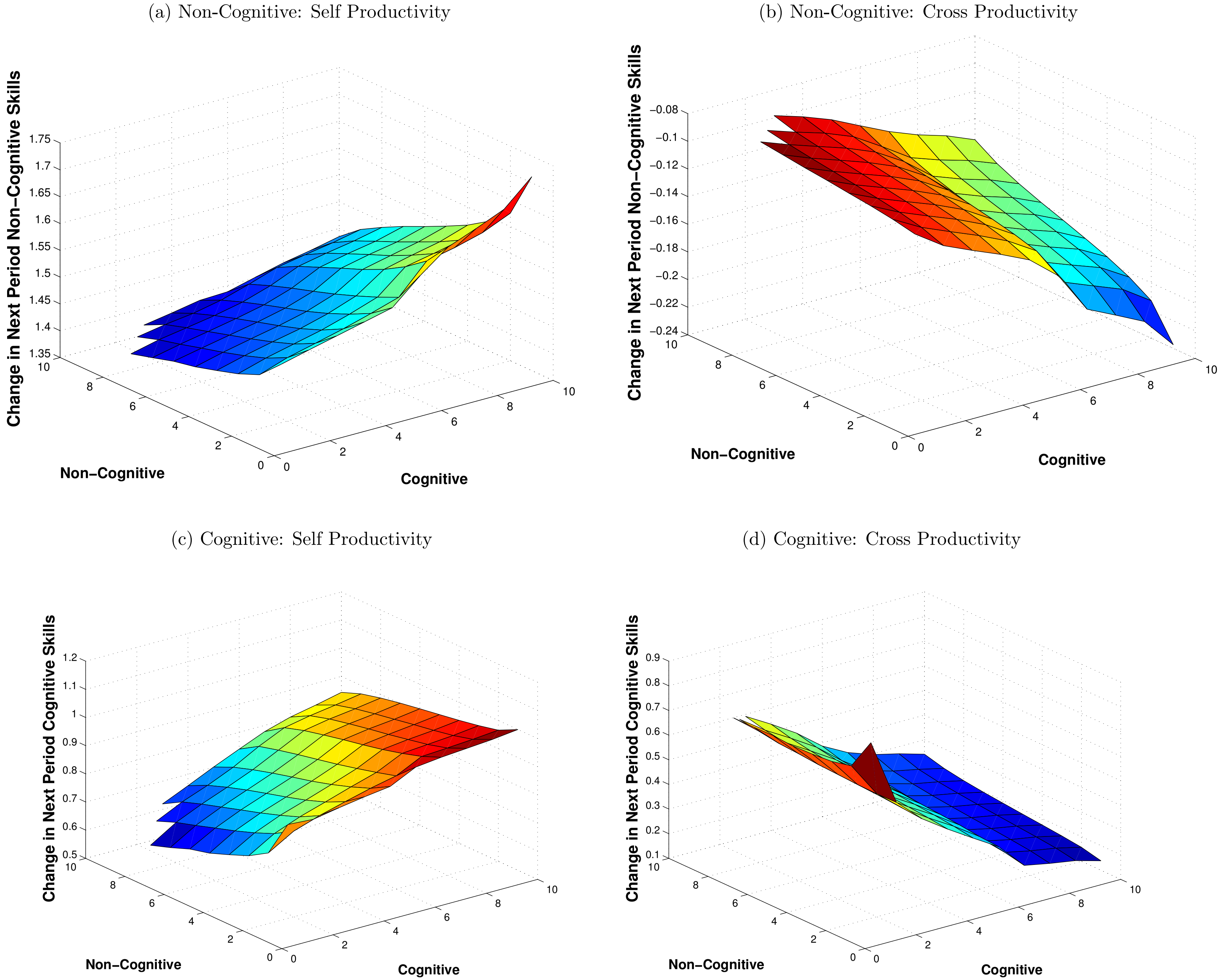


Figure 10: Self ($\partial\theta_{t+1}^S/\partial\theta_t^S$) and Cross ($\partial\theta_{t+1}^S/\partial\theta_t^{S'}$) Productivities ($t = 2$)



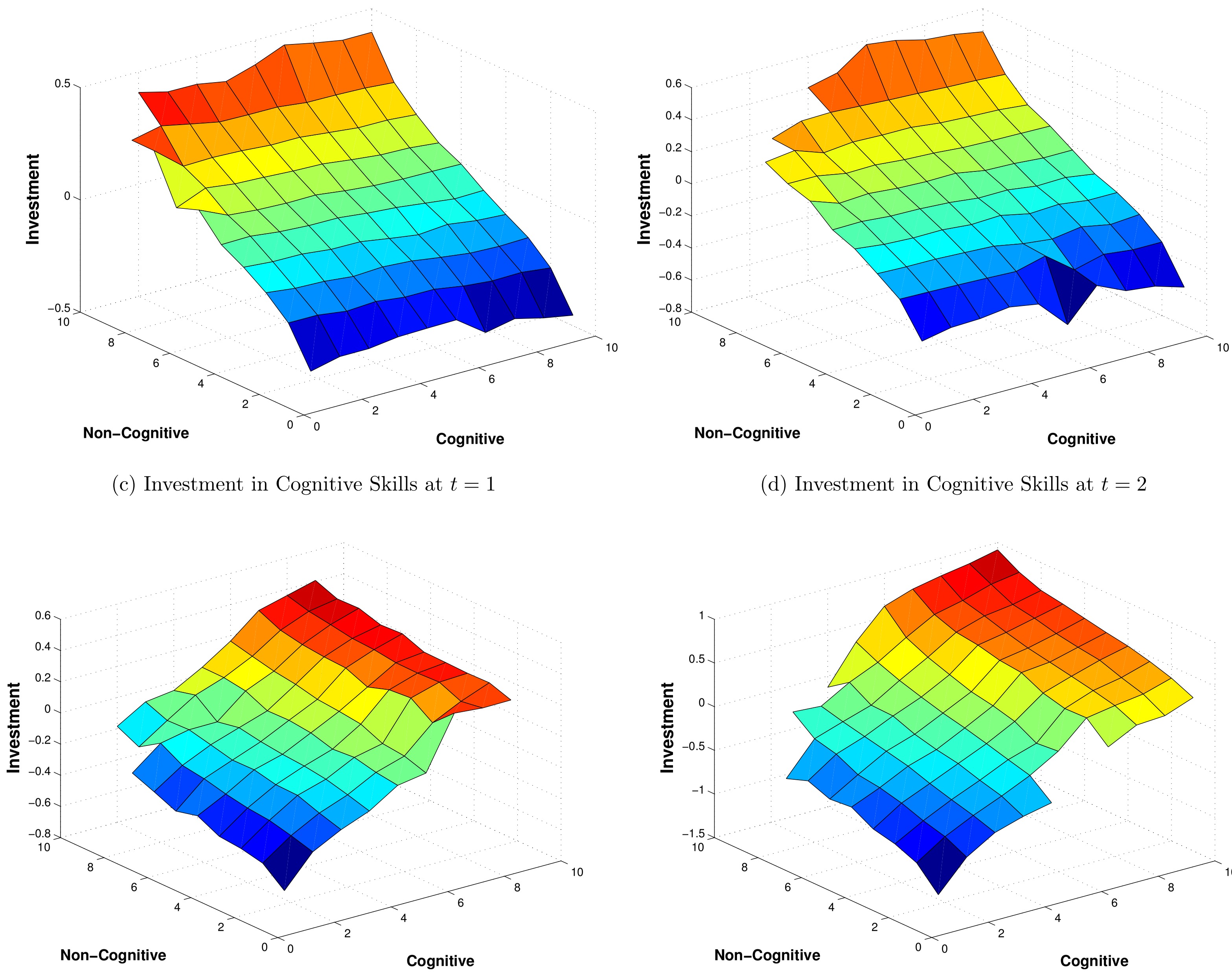
(a) Non-Cognitive: Self Productivity

(b) Non-Cognitive: Cross Productivity

(c) Cognitive: Self Productivity

(d) Cognitive: Cross Productivity

Figure 11: Skill Investment as a function of θ_t^{NC} and θ_t^{C}



(a) Investment in Non-Cognitive Skills at $t = 1$

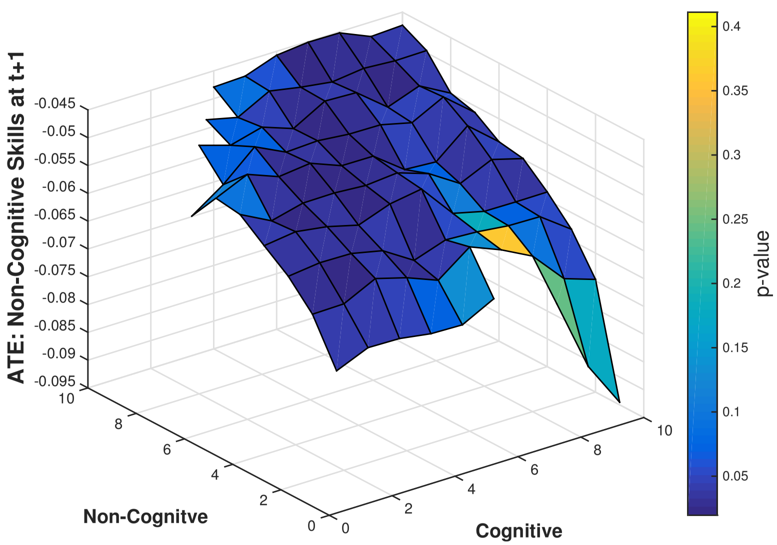
(b) Investment in Non-Cognitive Skills at $t = 2$

(c) Investment in Cognitive Skills at $t = 1$

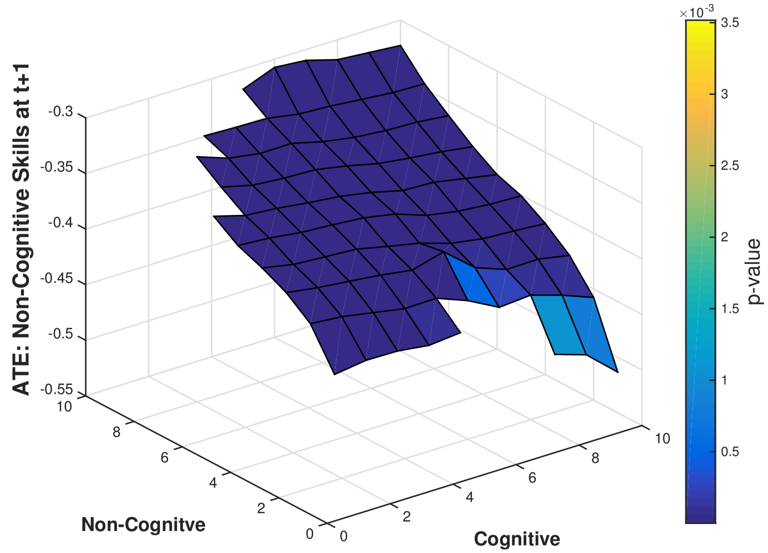
(d) Investment in Cognitive Skills at $t = 2$

Figure 12: $E\left[\theta^S_{t+1}|\theta^{NC}_t, \theta^N_t, M_{t+1}=1\right] - E\left[\theta^S_{t+1}|\theta^{NC}_t, \theta^N_t, M_{t+1}=0\right]$

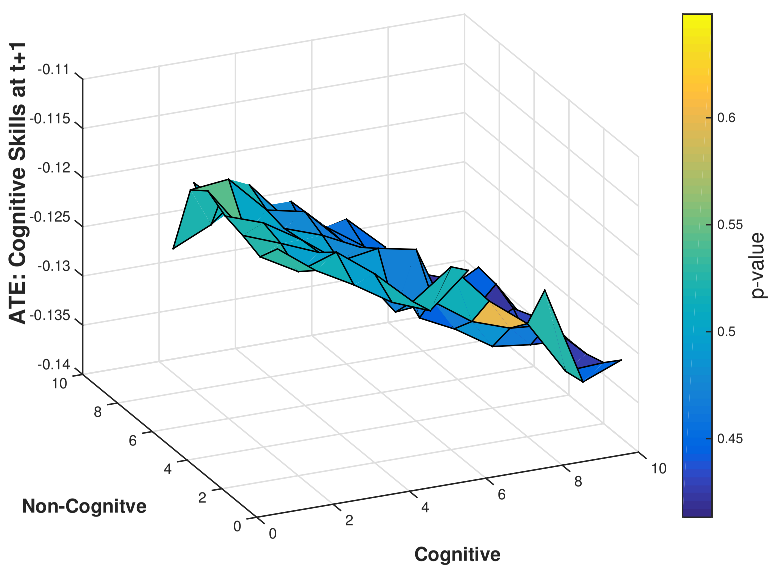
(a) Non-Cognitive $t=1$



(b) Non-Cognitive $t=2$



(c) Cognitive $t=1$



(d) Cognitive $t=2$

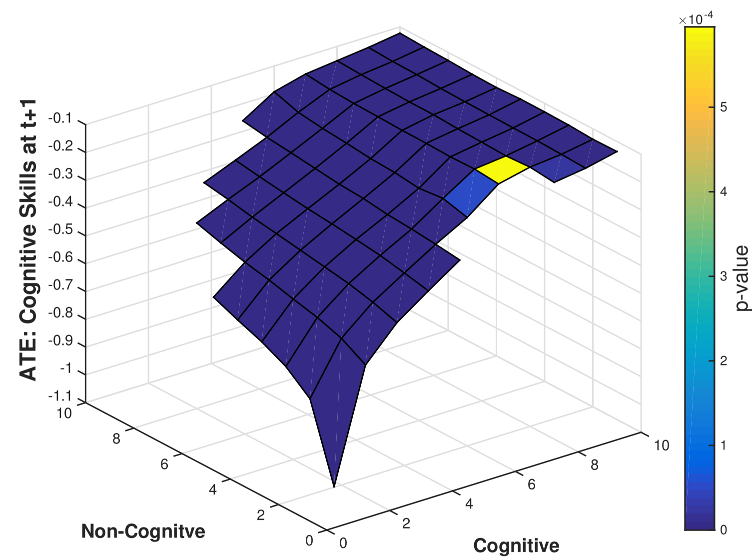
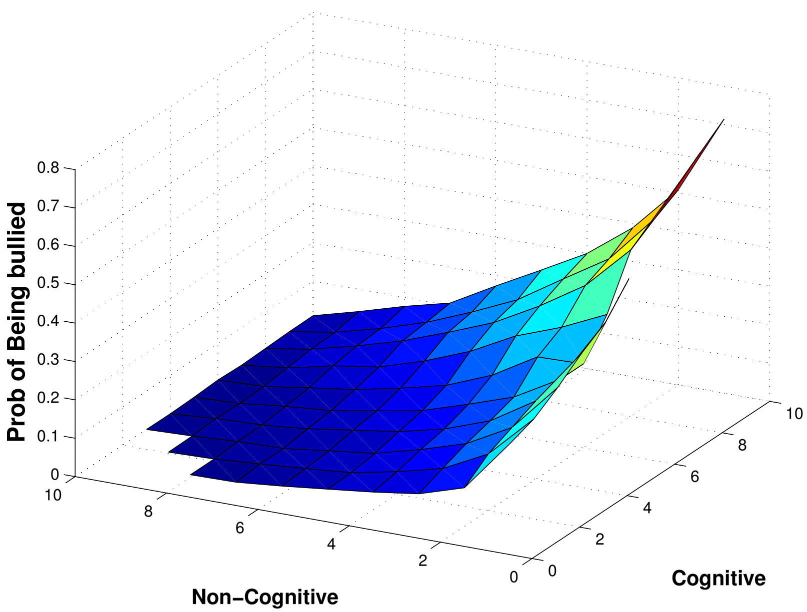


Figure 13: Probability of being bullied at $t = 2$ depending on bullied or not at $t = 1$

(a) Not bullied at $t = 1$



(b) Bullied at $t = 1$

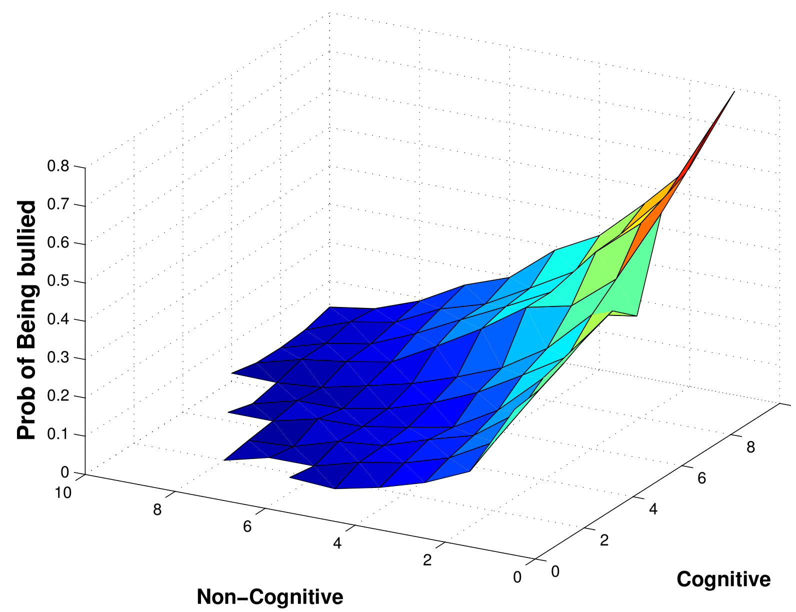
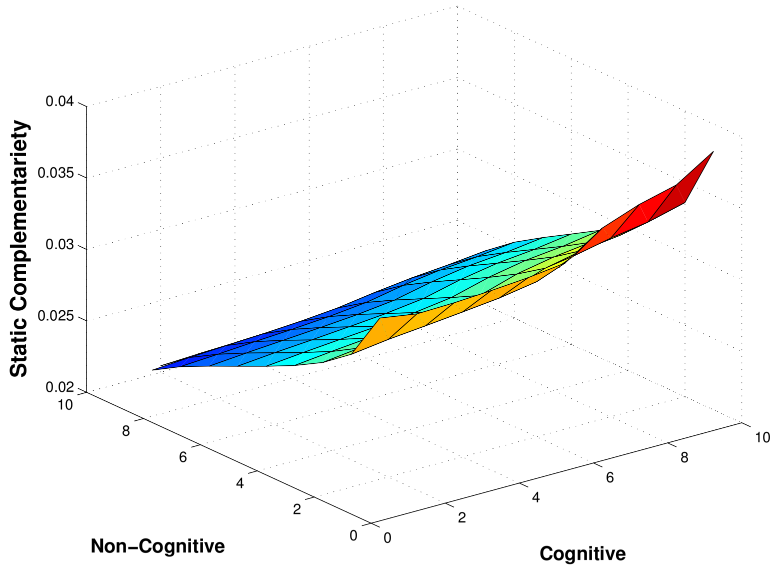
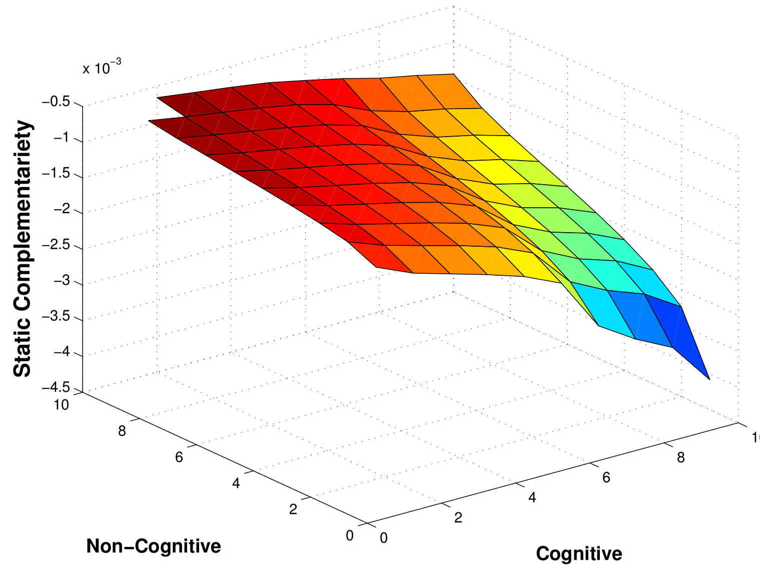


Figure 14: Static Complementariety at $t = 1$

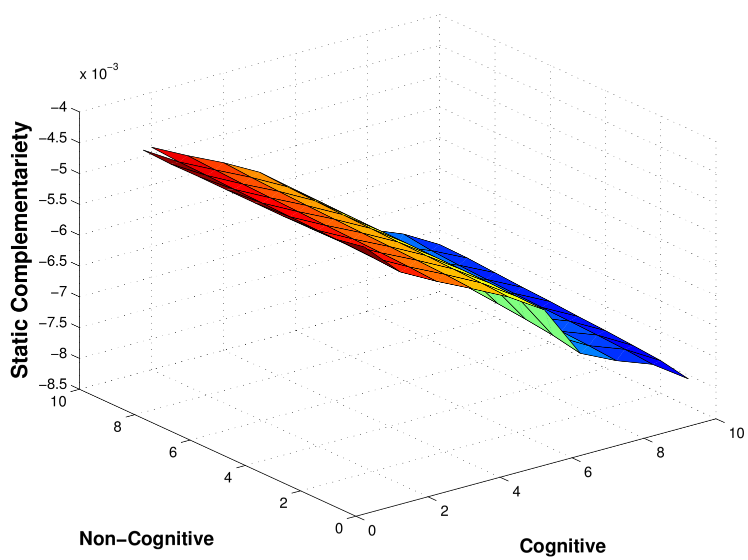
(a) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$



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



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



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

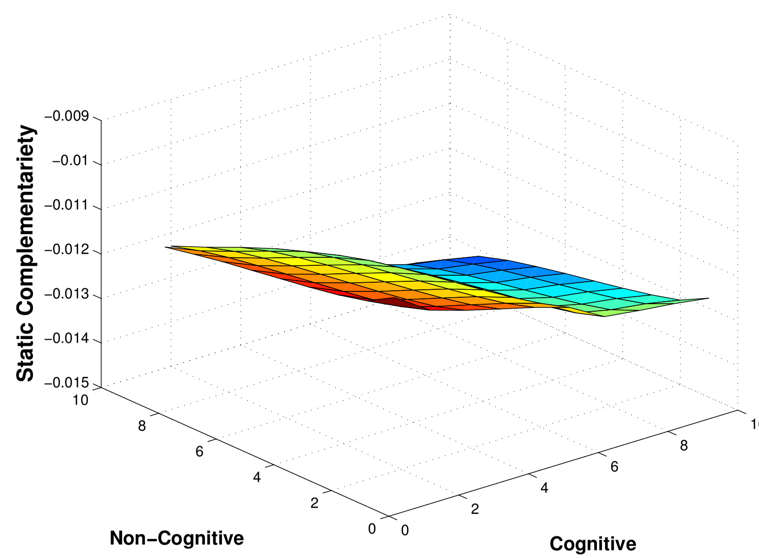
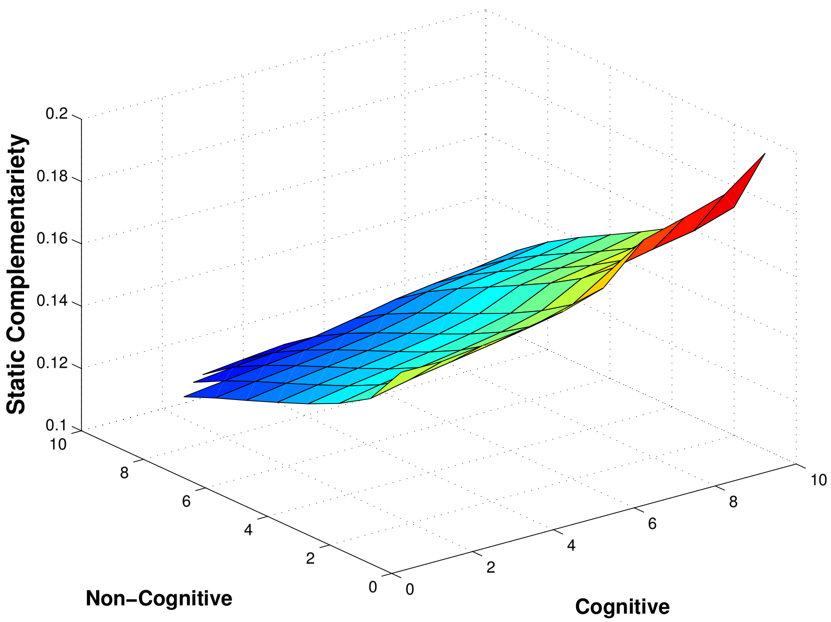
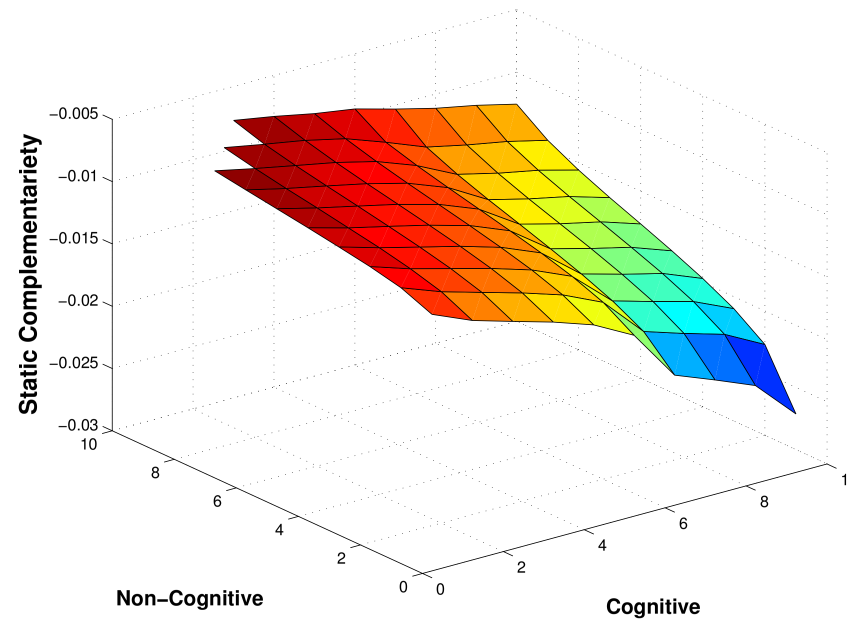


Figure 15: Static Complementariety at $t = 2$

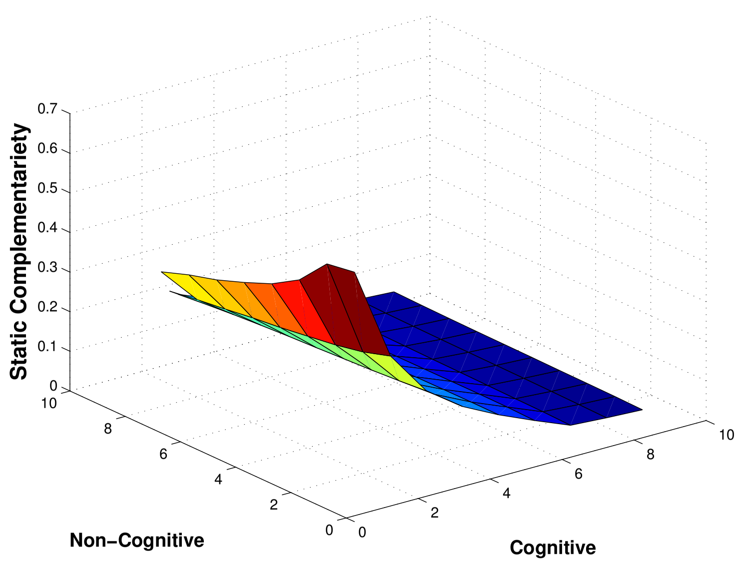
(a) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$



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



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



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

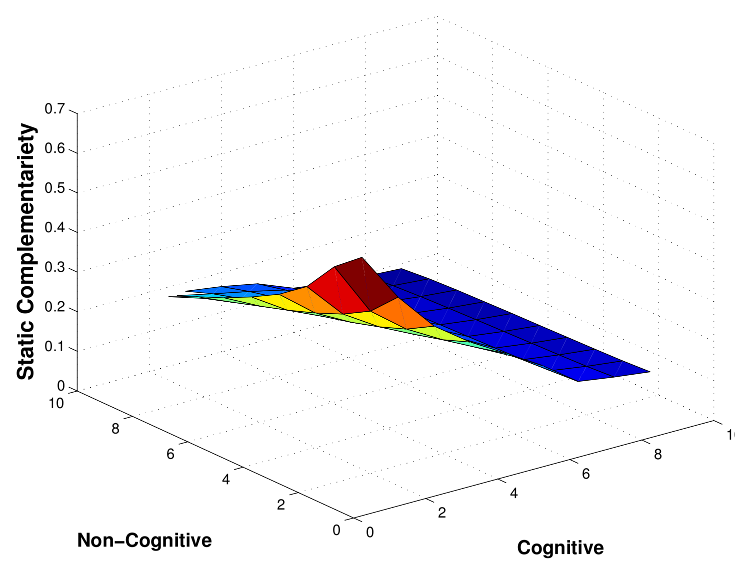


Figure 16: Dynamic Complementariety

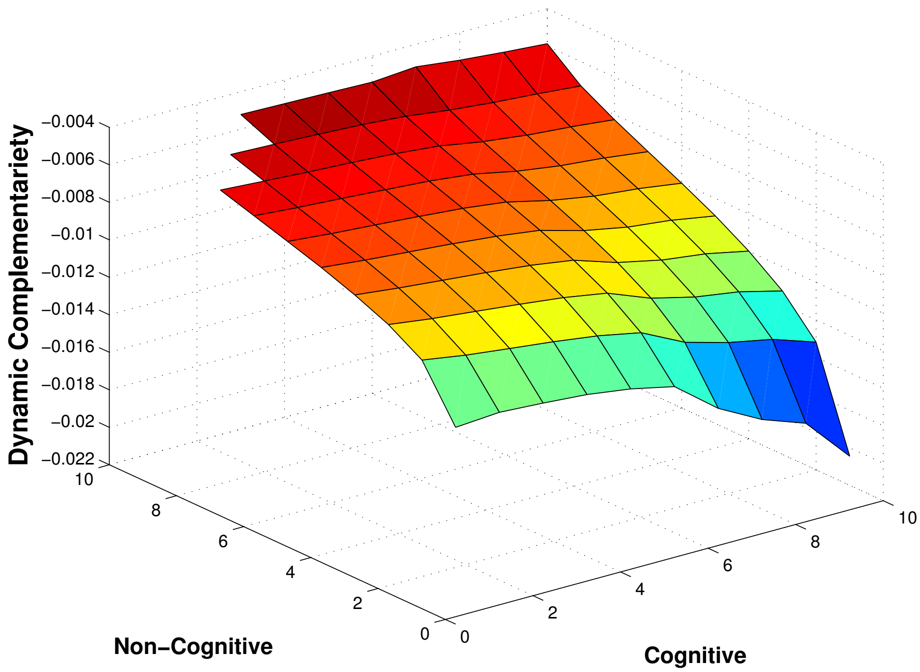
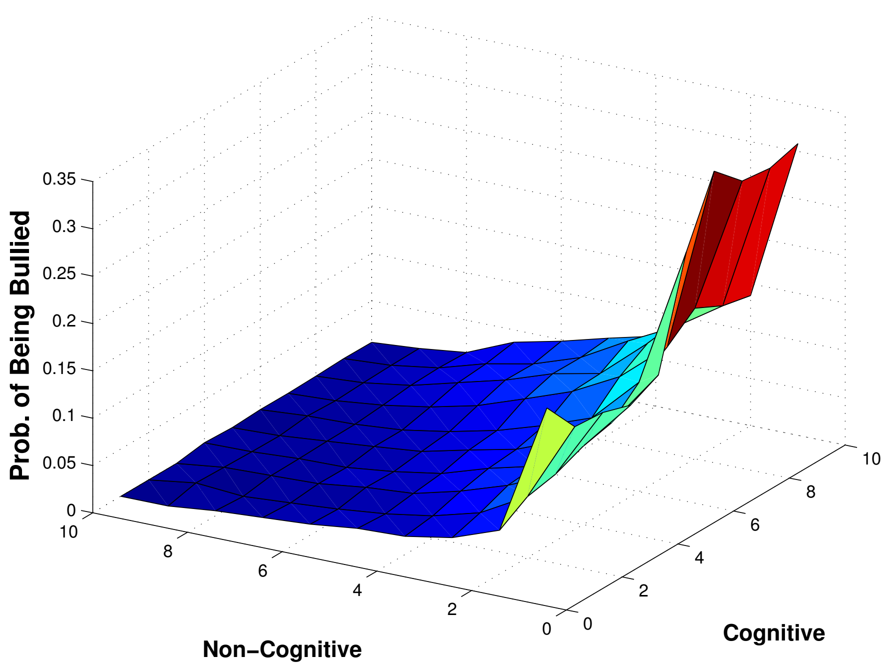


Figure 17: Classroom Allocation Simulations: Benchmark



Appendix

A Attrition Analysis

In this Appendix, I present some estimations regarding the observations lost due to attrition. First, Table A.1 shows that attrition in the first three waves —which used for the estimations done in this paper— is relatively low compared with similar surveys.

Table A.1: Attrition by Wave

	Attrit.	Attr. if Bull in $t = 1$	
Wave		No	Yes
1	.	.	.
2	7.5%	7.7%	7%
3	9.4%	9.1%	10.4%
4	9.5%	9.1%	11%
5	13.9%	13.7%	14.8%

In addition, Tables A.2 and A.3 show that there are few differences between those who leave the sample and those who stay. 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 or victimization. In Table A.3, I incorporate the unobservables (i.e., cognitive and non-cognitive skills) in the analysis. It shows that, consistent with the findings in Table A.2, the kids that leave the sample are low cognitive skilled wealthy kids with highly educated parents.

Table A.2: Difference in Observables at $t = 1$ of Attrited and Non-Attrited Observations

Variable	Mean Att	Mean Stay	Diff.
Age (months)	8.6346	8.9626	-.328
Male	.5019	.5	.0019
Older Siblings	.4559	.5452	-.0893*
Young Siblings	.6398	.6341	.0058
lnInc_pc	4.5632	4.3275	.2356*
Urban	.8659	.8676	-.0017
Lives: Both Parents	.9195	.9294	-.0099
Lives: Only Mother	.0383	.0332	.0051
Father Edu: 2yColl	.0728	.0678	.005
Father Edu: 4yColl	.295	.2974	-.0023
Father Edu: GS	.1341	.063	.0711*
Locus of Control	.0631	-.0052	.0682
Irresponsibility	-.0827	.0068	-.0895
Self-Esteem	.0068	-.0006	.0074
Language & SocStd	-.0907	.0074	-.0981
Math & Science	-.1457	.0119	-.1576*
Yearly Test	-.108	.009	-.117*
Bullied	.2107	.2262	-.0154
Bully	.2759	.2437	.0321

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

Table A.3: Probability of Staying from $t = 1$ to $t = 2$

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

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

B Proofs

B.1 Proof Theorem 1

Organize the adjunct measurement system (4) 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, we can partition the measurement system in two blocks

$$\begin{bmatrix} \mathbf{T}^A \\ \mathbf{T}^{A,B} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_T \beta^T + \alpha^{(A)A}\theta^A + \mathbf{e}^T \\ \mathbf{X}_T \beta^T + \alpha^{(A,B)A}\theta^A + \alpha^{(A,B)B}\theta^B + \mathbf{e}^T \end{bmatrix}$$

Assuming that the latent factor A has non-degenerate distributions

$$COV\left(T_h^A, T_k^A\right) = \alpha_h^{(A)A}\alpha_k^{(A)A}\sigma^2_{\theta^A} \qquad \text{for } h,k = 1,\ldots,L_A; \quad h \neq k$$

and

$$COV\left(T_h^A, T_l^A\right) = \alpha_h^{(A)A}\alpha_l^{(A)A}\sigma^2_{\theta^A} \qquad \text{for } h,l = 1,\ldots,L_A; \quad h \neq l$$

Hence,

$$\frac{COV\left(T_h^A, T_k^A\right)}{COV\left(T_h^A, T_l^A\right)} = \frac{\alpha_k^{(A)A}}{\alpha_l^{(A)A}}$$

Therefore, $L_A - 1$ factor loadings are identified up to one normalization. Also, assuming with out loss of generality that the normalized loading is that of equation l we have that

$$\sigma^2_{\theta^A} = \frac{COV\left(T_h^A, T_l^A\right) COV\left(T_k^A, T_l^A\right)}{COV\left(T_h^A, T_k^A\right)}$$

Now turning to the second block of the measurement system and assuming that latent factor B has a non-degenerate distribution, I have that

$$COV\left(T_m^B, T_n^B\right) = \alpha_m^{(A,B)A}\alpha_n^{(A,B)A}\sigma_{\theta^A}^2 + \alpha_m^{(A,B)B}\alpha_n^{(A,B)B}\sigma_{\theta^B}^2 + \left(\alpha_m^{(A,B)A}\alpha_n^{(A,B)B} + \alpha_m^{(A,B)B}\alpha_n^{(A,B)A}\right)\sigma_{\theta^A,\theta^B} \tag{18}$$

for $m, n = 1, \ldots, L_B$ and $m \neq n$, and

$$COV\left(T_m^B, T_l^A\right) = \alpha_m^{(A,B)A}\alpha_l^{(A)A}\sigma_{\theta^A}^2 + \alpha_m^{(A,B)A}\alpha_l^{(A)A}\sigma_{\theta^A,\theta^B} \tag{19}$$

for $m = 1, \ldots, L_B$ and $l = 1, \ldots, L_A$. It is easy to see that this second block of the measurement system is not identified as it has $2L_B + 2$ unknowns, while it has only $2L_B$ pieces of relevant information. That is, I have two loadings per measure plus $\sigma_{\theta^B}^2$ and $\sigma_{\theta^A,\theta^B}$ to identify, and I have two covariances per each L_B tests: one with another second block test and another one with one the L_A test (preferably, the one that has the normalized loading).

Now, let me reduce the number of unknowns by normalizing one of the second block loadings (i.e., $\alpha_o^{(A,B)B} = 1$ for $o = \{1, .., L_B\}$) and assuming one of the second block measures is only affected by the second factor (i.e., $\alpha_o^{(A,B)A} = 0$), then

$$COV\left(T_m^B, T_o^B\right) = \alpha_m^{(A,B)B}\sigma_{\theta^B}^2 + \alpha_m^{(A,B)A}\sigma_{\theta^A,\theta^B} \tag{20}$$

Using the normalization of $\alpha_l^{(A)A} = 1$ I reported above, (19) becomes

$$COV\left(T_m^B, T_l^A\right) = \alpha_m^{(A,B)A}\sigma_{\theta^A}^2 + \alpha_m^{(A,B)B}\sigma_{\theta^A,\theta^B} \tag{21}$$

Furthermore, $COV\left(T_l^A, T_o^B\right) = \sigma_{\theta^A,\theta^B}$. Then, using (20) and (21), I can identify $\alpha_m^{(A,B)A}$ and $\alpha_m^{(A,B)B}$ as a function of measurement covariances and $\sigma_{\theta^B}^2$, which I can later identify using equation

(18).

B.2 Proof Theorem 2

To simplify exposition let us reduce the number of parameters to two, and rename ρ as γ_2. Hence, the parameter set is $\mathbf{\Gamma} = \{\gamma_1, \gamma_2\}$. Assuming separability of the error term function to estimate is

$$\theta_{t+1} = h(\theta_t, \mathbf{\Gamma}) = \left(\gamma_1 \theta_{t,A}^{\gamma_2} + (1-\gamma_1)\,\theta_{t,B}^{\gamma_2}\right)^{1/\gamma_2} + \varepsilon$$

I linearize the CES function using a Taylor approximation around $\mathbf{\Gamma}^0$ in order to rely on the results of the linearized regression model described in Green (2000). That is, if $\theta_k^0 = \partial h(\theta_t, \mathbf{\Gamma}) / \partial \gamma_k^0$

$$h(\theta_t, \mathbf{\Gamma}) \backsimeq h\left(\theta_t, \mathbf{\Gamma}^0\right) - \gamma_1^0 \theta_1^0 - \gamma_2^0 \theta_2^0 + \gamma_1 \theta_1^0 + \gamma_2 \theta_2^0$$

or if I stack the variables into matrices

$$\theta_{t+1} \backsimeq h\left(\theta_t, \mathbf{\Gamma}^0\right) - \gamma^0 \theta^0 + \gamma \theta^0 + \varepsilon$$

Defining ε^0 as the error term that contains the true disturbance ε and the deviation that arises due to the Taylor approximation, and $\theta_{t+1}^0 = \theta_{t+1} - h(\theta_t, \mathbf{\Gamma}^0) + \gamma^0 \theta^0$ then I have

$$\theta_{t+1}^0 = \gamma \theta^0 + \varepsilon^0$$

which can be estimated using least squares. Given that I am particularly interested in γ_2, let me face the estimation procedure using the partitioned regression.

$$\theta_{t+1}^0 = \begin{bmatrix} \theta_1^0 & \theta_2^0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \varepsilon^0$$

From the Frisch-Waugh theorem we now that $\hat{\gamma}_2$ can be obtained by regressing the residuals from the regression between θ^0_{t+1} and θ^0_1 on the residuals from the regression between θ^0_2 and θ^0_1. That is,

$$\hat{\gamma}_2 = \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 M_{\theta^0_1} \theta^0_{t+1}$$

where $M_{\theta^0_1} = I - \theta^0_1 \left(\theta^{0\prime}_1 \theta^0_1\right)^{-1} \theta^{0\prime}_1$ is the annihilator matrix of the regression of θ^0_{t+1} on θ^0_1. Hence, I can check the bias in $\hat{\gamma}_2$

$$\begin{aligned}
E\left[\hat{\gamma}_2 | \theta^0_1, \theta^0_2\right] &= E\left[\left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 M_{\theta^0_1} \left[\theta^0_1 \gamma_1 + \theta^0_2 \gamma_2 + \varepsilon^0\right]\right] \\
&= \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 \left[M_{\theta^0_1} \left[\theta^0_1 \gamma_1 + \theta^0_2 \gamma_2\right] + M_{\theta^0_1} E\left[\varepsilon^0\right]\right] \\
&= \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 \left[M_{\theta^0_1} \theta^0_2 \gamma_2 + M_{\theta^0_1} E\left[\varepsilon^0\right]\right] \\
&= \gamma_2 + \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 M_{\theta^0_1} E\left[\varepsilon^0\right]
\end{aligned}$$

Therefore the, bias in $\hat{\gamma}_2$ depends on the mean of ε^0 which in turn depends on the mean of θ_{t+1} trough ε. Hence, the only way $E\left[\hat{\gamma}_2\right] = \gamma_2$ is if $E\left[\theta_{t+1}\right] = 0$.

C Information Used to create Non-cognitive score

In the case of locus of control, we added 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 we added:

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 we created the irresponsibility index by adding:

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

D Information used to create the measures of investment in non-cognitive skills

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 following variables:

- I frequently see my parents verbally abuse each other
- I frequently see one of my parents beat the other one
- I am often verbally abused by parents
- I am often severely beaten by parents

Table D.1: Descriptive Statistics of Investment Indexes

	Parental Abuse		Parental Control		Parental Harmony	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
All	7.1641	3.157	0.002	0.997	0	0.999
Males	7.1627	3.094	-0.118	0.983	-0.086	0.963
Females	7.1654	3.219	0.124	0.997	0.085	1.025
Attending College*	6.8914	3.008	0.038	1.012	0.036	0.969
Not Attending College*	7.5211	3.309	-0.043	0.098	-0.048	1.034

* Sample limited to wave 6

The parental control index is created by aggregating:

- When I go out, my parents usually know where I am
- When I go out, my parents usually know whom I am with
- When I go out, my parents usually know what I do
- When I go out, my parents usually know when I return

Finally, the parental harmony index is created using the following variables:

- My parents and I try to spend much time together
- My parents always treat me with love and affection
- My parents and I understand each other well
- My parents and I candidly talk about everything
- I frequently talk about my thoughts and what I experience away from home with my parents
- My parents and I have frequent conversations

E Outcomes Analysis at age 16

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

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

Its purpose is to capture the effect of skills on more tangible outcomes, and in that way have a better picture about how the skills lost to bullying hurt the development of successful lives. See Sarzosa and Urzua (2013b) for a detailed explanation on how the outcome measures were constructed.

Table E.1: Effect of unobserved heterogeneity at age 16

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

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