Essays on the Economics of Education

by

Jiemin Xu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)

at the
University of Wisconsin-Madison
2025

Date of Final Oral Exam: 04/28/2025
The dissertation is approved by the following members of the Final Oral Committee:

Christopher R Taber, Professor, Economics
Chao Fu, Professor, Economics
Matthew Wiswall, Professor, Economics
Dayin Zhang, Assistant Professor, Real Estate and Urban Land Economics

Essays on the Economics of Education

Jiemin Xu

Abstract

The first chapter studies children as active decision-makers in their own skill development during adolescence and analyzes how parents influence children's choices through strategic interaction. Using unique panel data on intra-household interactions from South Korea, I investigate a non-pecuniary and indirect form of parental involvement in education - parents' discussions about performance with children. Exploiting within-individual variation, I find that regular parent-child performance discussion increases the child's time paying attention in class by 13 percentage points. I incorporate this interaction into a dynamic model of skill formation where performance discussion is costly to parents but increases child effort. Parents can enhance skill development through this novel indirect channel in addition to a standard direct channel of educational investment. Shutting down the channel of performance discussion decreases skill accumulation by 0.09 standard deviation in four years, equivalent to a \$5,140 reduction in educational investment. By disentangling the determinants of performance discussion in the model, I find that differential costs rather than benefits result in more performance discussion in high-socioeconomic-status households. Eliminating the cost difference slows the expansion of the skill gap between high- and low-socioeconomic groups in adolescence by 7.9 percent.

The second chapter studies the educational challenges faced by rural-to-urban migrant children in China. I document that compared to urban local students, migrant children have lower cognitive skills and receive fewer educational investments from both schools (restricted access to high-quality public schools) and parents. I estimate a generalized skill production function and find negative complementarity between initial skill and school and parental investments. This indicates that increasing investments in migrant children (who start with lower skills) can effectively close the skill gap between migrant and local students. To show how eliminating the disparities in school and parental inputs can reduce the migrant-local skill gap, I conduct two counterfactual exercises. I find that 1) if migrant children have the same access to schools as local students, the skill gap is reduced by 42 percent, and 2) providing migrant children with the average parental investment received by local students closes the skill gap by 17 percent.

Dedication

To my family—my mom, dad, sister, and fiancé—for their unconditional love and support. And to my advisors, for their invaluable guidance and mentorship.

Declaration

I hereby declare that this dissertation is the result of my own work and research, conducted under the supervision of Christopher Taber, Chao Fu, and Matthew Wiswall. All sources of information and data have been duly acknowledged. This dissertation has not been submitted for any other degree or diploma at any other institution.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisors, Christopher Taber, Chao Fu, and Matthew Wiswall. Your strong support, insightful advice, and constant encouragement have been the cornerstone of my PhD journey. Chris, thank you for helping me think deeply about fundamental economic questions and for challenging me to tackle endogeneity issues. Chao, your guidance has been instrumental in deepening my understanding and expertise in structural modeling. Matt, your works have inspired my exploration in the area of parental influence and child development. I am incredibly fortunate to have had the opportunity to learn from each of you.

I would also like to extend my heartfelt thanks to other faculty members in the Department of Economics at the University of Wisconsin-Madison. Your valuable feedback and constructive criticism during seminars and discussions have significantly contributed to the development of my dissertation. Special thanks to Naoki Aizawa, Jesse Gregory, and John Kennan for your insightful comments on structural modeling; to Fernanda Rojas-Ampuero and Jeffrey Smith for your thoughtful advice on reduced-form analysis; and to Corina Mommaerts and Martin O'Connell for your brainstorming with me on motivating and structuring my paper.

My sincere appreciation goes to my peers and colleagues in the empirical microeconomics group of the Department of Economics at the University of Wisconsin-Madison. The collaborative environment and intellectual exchanges have been instrumental in refining my ideas and approaches. I appreciate the camaraderie and support from Qin Lin and Xian Wu who have been wonderful companions throughout these years.

I am deeply thankful to my friends, who have been my pillars of strength. Your encouragement, understanding, and empathy have been invaluable. Special thanks to Congyan Han for supporting me during challenging times, and to Jinglin Yang and Huaxia Xiong for always being there for me.

To my family—my mom, dad, sister, and fiancé—words cannot express how grateful I am for your unconditional love and support. Your belief in me has been my greatest motivation. Mom and Dad, thank you for embracing me as I am and raising me to be a brave and independent adventurer who never stops pushing the limits. To Yixuan, thank you for always being my biggest cheerleader and sweetest little sister, trusting me and standing by my side under all circumstances. And to Yunlu, thank you for encouraging me to stay true to myself during times of uncertainty and self-doubt in my early 20s. Your endless empathy and love have been my anchor throughout this journey.

Thank you all for being a part of this adventure. Your contributions have made this achievement possible.

Contents

List of Figures

List of Tables

Chapter 1

Parent-Child Strategic Interaction and Skill Development in Adolescence

1.1 Introduction

Parental actions have a significant impact on child development, and differential parenting decisions contribute to disparities in child outcomes (Heckman and Mosso, 2014). While the existing literature has recognized the importance of parental investments in skill production during early childhood,[1] This paper goes beyond direct parental impact from educational inputs to explore additional roles parents play in child development during later periods. Specifically, this paper attempts to study adolescents' decisions about their own educational investment and investigate the indirect channel that parents influence skill development by affecting child decision-making. The investigation is driven by two developmental phenomena in adolescence.[2] First, adolescents go through a transition from

[1] The majority of this literature has focused on one or both of two typical forms of parental investments - expenditure on educational goods and parental time with the child. Heckman and Mosso (2014) is an extensive summary of the evidence on the importance of the early home environment in shaping skills.

[2] According to the World Health Organization (WHO), adolescence is the phase of life between childhood and adulthood, from ages 10 to 19.

passive dependents on parents to active decision-makers. Despite this fact, the child development literature has often ignored child independence in decision-making, and little is known about how adolescents make learning decisions. Second, adolescents experience rapid development in advanced cognitive skills that are vital for success in the labor market. The major form of learning—schooling—is not directly supervised by parents, so parents are more likely to influence child schooling behaviors through other indirect forms of involvement.[3]

This paper focuses on an indirect form of parental involvement in schooling, parents' discussion about performance, through which parents communicate the value of good performance to their children. Using a unique panel dataset on parent-child interaction for adolescents in South Korea - the Korean Education Longitudinal Study (KELS),[4] I find that performance discussion is a prevalent parenting practice: around 55 percent of parents have such discussions with their children regularly. Exploiting within-child variation, I provide causal evidence that parents' discussion about performance increases the child's time paying attention in class by 12.78 percentage points. Moreover, I find that college-educated parents are more likely to engage in performance discussions, which may have implications for skill disparities across socioeconomic groups.

Motivated by the empirical evidence, I build a dynamic model of parent-child strategic interaction within the framework of child cognitive development to characterize 1) the impact of performance discussion on child effort choices and its implication for skill development and 2) the determinants of parents' performance discussion. The model has four main features. First, in contrast to classical child development models that treat children as passive recipients of investments,[5] the model acknowledges children as active

[3]The existing literature has documented one important source of parental influence on schooling, that is, parents decide which school their child attends. This paper takes school choices as given and explores how parental involvement in schooling influences child learning behaviors in school on a daily basis.

[4]KELS tracks a nationally representative sample of children in South Korea from primary to secondary education in 2013-2020.

[5]Currie and Almond (2011) is an extensive survey of this literature. The majority of this literature focuses on preschool children, to whom parents are dominant actors and have direct control over the development process. Given the dependence of young children on their parents, it is reasonable to assume children are passive receivers. However, child cognitive development does not end after early childhood, and how school-age children behave as active decision-makers remains largely unknown.

decision-makers and allows them to determine their effort in learning. Second, parents can indirectly influence child skill development through a novel channel of performance discussion that affects child effort choices, in addition to the standard direct channel of parental educational investments. The modeling of an indirect channel builds on the theoretical works that formalize parental strategic influence on children's educational choices through pecuniary incentives or corporal punishment in structural models (Akabayashi, 1996; Weinberg, 2001; Doepke and Zilibotti, 2017). The model considers a different non-pecuniary parenting strategy that emphasizes the value of good performance to influence children's behaviors. Third, skill development is accumulative, and the skill technology reflects diminishing returns of educational inputs and positive complementarity among inputs. Fourth, the model allows for correlated heterogeneity in unobserved preferences and abilities to account for the endogeneity in choices of educational inputs and parent-child interactions. In particular, performance discussion hinges on the trade-off between cost and benefit, and individual-specific preferences and abilities allow for heterogeneous costs and benefits, resulting in differential choices of performance discussion.

The uniquely rich KELS panel data on parent-child performance discussion, educational inputs, and cognitive test scores permits the identification of individual-specific learning ability and preference parameters, which address the individual-level endogeneity in behaviors. To quantify the contributions of the performance discussion channel to average skill development and skill inequality, I estimate the model in two steps. First, I exploit within-child variations in observed educational inputs to estimate the skill technology, assuming individual-specific time-invariant learning efficiency (or total productivity factor) and exogenous temporary production shocks.[6] Individual-specific TFPs capture all time-invariant observed and unobserved heterogeneity in child and family characteristics,[7] as well as any fixed omitted inputs. Replacing individual-specific TFPs with rich

[6]Conditional on individual-specific TFPs and previous skill, temporary production shocks are assumed exogenous to educational investments. This assumption is weaker than the common practice in the existing literature on the technology of skill formation or education production function that assumes exogenous production shock conditional on previous skills.

[7]For example, observed demographics commonly included as controls in production function (child gender, number of siblings, parental education level, etc.) and unobserved traits like child ability to learn.

demographic controls and school fixed effects creates positive biases in the estimated productivity parameters, suggesting positive correlations between observed inputs and unobserved individual heterogeneity not proxied by demographics and school fixed effects. This finding serves as the primary motivation for exploiting within-child variations. In the second step, conditional on production parameters, household preference parameters are revealed by observed household choices and estimated using the Simulated Method of Moments.

The estimated skill technology shows that skill is more elastic in child effort than in parental investment. A one-percent increase in child effort increases skill by 0.23 percent, while a one-percent increase in parental investment raises it by 0.11 percent. The estimated structural model finds significant correlations among unobserved heterogeneity in household characteristics that account for endogeneity in household choices of educational inputs and performance discussion. First, the correlation between child learning efficiency and parents' consumption preference (children's leisure preference) explains the endogeneity of parental investment (child effort) to TFPs highlighted in the estimation of skill technology. Moreover, the estimated model shows that variations in cost and benefit are two sources of heterogeneity in performance discussion. Parents' discussion cost is correlated with child learning efficiency and child leisure preference, and parents' benefit from discussion increases in child leisure preference. Taken together, the model confirms that parents' choice of performance discussion depends on unobserved child characteristics. Lastly, parents' preference for consumption is positively correlated with their performance discussion cost, resulting in an endogenous positive correlation between educational investment and performance discussion. Overall, the estimated model provides a good fit for the data patterns of child effort, parental educational investment, and performance discussion.

Based on the estimated model, I quantify the importance of parents' performance discussion to child effort and its implication for skill formation. First, by exploiting within-child variation, the model confirms a significant and causal linkage between performance discussion and child effort.[8] I find that a substantial portion of child effort (10.4-12.1 percent on average) can be attributed to performance discussion. Second, I generate the counterfactual path of skill development when the channel of performance discussion shuts down and compare it to the path predicted in the baseline model. Shutting down the performance discussion channel decreases skill accumulation by 0.09 standard deviation in four years, equivalent to an approximately $5,140 reduction in parental educational investment. Therefore, ignoring indirect parental impact from performance discussion will substantially understate parental influence on skill development.

In the baseline model, not all parents choose to discuss performance, which suggests that promoting performance discussion in the population can improve the skill formation of children whose parents never or rarely discuss performance. I simulate household choices and generate the skill path in a counterfactual model that facilitates performance discussion for all parents. I find that average child effort increases by 6.65-8.60 percent from the baseline, resulting in a 0.040 standard deviation increase in skill accumulation over four years. To achieve the same increase in skill, it is equivalent to increasing parental educational investment by $3,922. Quantifying the impact of promoting performance discussion forms the basis of forecasting the effectiveness of policy interventions encouraging parental involvement in schooling to improve skill.

Having established the significance of performance discussion in shaping average skill formation, I explore its potential role as a source of skill disparity. The investigation is guided by the model, which suggests that performance discussion may influence skill distribution through two mechanisms. First, parents face heterogeneous performance discussion costs and benefits, leading to differential probabilities of performance discussion. Second, the effect of performance discussion on child effort varies across population by child characteristics. In particular, I find that low-socioeconomic-status[9] parents receive higher discussion benefits because performance discussion has a larger effect on their chil-

[8] Despite several sources of potential downward bias in estimation, I find a significantly positive effect of performance discussion on effort, which serves as a lower bound of the actual effect. I discuss the potential biases in estimation in the appendix.

[9] I measure socioeconomic status (SES) by maternal education and define a household with a college-educated mother as high SES.

dren's skill accumulation. However, they face even higher discussion costs, leading to a lower likelihood of performance discussion. This exacerbates the skill gap between high- and low-SES children during adolescence. From fifth to ninth grade, the skill gap widens by 0.230 standard deviation. Eliminating the difference in performance discussion cost across high- and low-socioeconomic groups closes this expansion of the skill gap by 0.018 standard deviation or 7.9 percent.

Related Literature

This paper draws on and contributes to the growing theoretical and empirical analyses of parent-child strategic interaction. Among the first static models of parent-child interaction, Akabayashi (1996) formalizes the parent-child relationship as a principal-agent problem where parents provide incentives (reward and/or punishment) to influence child learning effort and Weinberg (2001) further investigates how parenting practices vary with income and argues that the limited ability of low-income parents to offer pecuniary incentive leads to their increased reliance on non-pecuniary mechanisms such as corporal punishment. More recent works in Doepke and Zilibotti (2017, 2019) provide theoretical insights about parenting styles that classify parent-child interactions into two types, constraining children's choice sets and molding children's preferences, which inspire how this paper models the two channels of parental influence. Only a few papers have attempted to analyze the dynamic game between parents and children empirically. For example, motivated by the empirical association between pecuniary incentives from parents and child self-investment, Del Boca et al. (2023) conjectures the mechanisms and the impact of this parenting strategy on child outcomes in a structural model of parent-child interaction. Relative to this strand of literature, the unique panel data on parent-child performance discussion allows me to provide causal estimates of the effect of a different non-pecuniary parenting strategy on child behaviors by exploiting within-child variation. I establish a novel channel of parental influence - parents enhance child skill development by affecting children's choices with a "soft" approach of communicating the value of good performance to children.

This paper extends the emerging literature on child effort choices in the context of education in two aspects. First, the focus on child in-class effort in this paper complements the analyses of child after-school study time (e.g., homework, reading, etc.) in Cooper et al. (2006) and Del Boca et al. (2017) to enhance the appreciation of the importance of child self-investments to skill formation. Broadly speaking, this paper and other works on the productivity of child self-investments expand on the extensive literature on the technology of skill formation that has centered on parental and public investment in children (Todd and Wolpin, 2003; Cunha and Heckman, 2007; Todd and Wolpin, 2007; Cunha et al., 2010; Agostinelli and Wiswall, 2025; Attanasio et al., 2020).[10] Furthermore, this paper examines the role of parents in child decision-making and improves the understanding of what factors affect child effort choices along with other recent papers that investigate the impact from teachers (Todd and Wolpin, 2018), peers (Conley et al., 2024), and researchers (Cotton et al., 2020).

This paper builds on and goes beyond the literature on structural modeling of parental investment in education (Cunha et al., 2013; Del Boca et al., 2014; Lee and Seshadri, 2019; Caucutt and Lochner, 2020; Bolt et al., 2023). These studies integrate the skill technology into a structural model of parental behavior to estimate the impact of parental investment on skill development. Explicitly modeling the mechanisms of choices helps explain the observed heterogeneity in parental investment and enables formal counterfactual analysis of policy interventions that alter parental behaviors. In the same spirit, this paper incorporates the analysis of parent-child strategic interaction into the framework of dynamic skill accumulation to fill the gap of ignoring child decision-making in those studies. By treating the child as an active decision-maker in educational investment and exploring parents' impact on child effort choices, I add an indirect channel of parental influence on skill development to the direct channel of parental educational investment and benchmark the relative importance of the indirect parental influence with the direct channel.

[10] This literature establishes several important features of the technology of skill formation: a) multiple periods in the life cycle in the formation of skills, b) multiple forms of investment.

Section 2 presents the dynamic model of parent-child strategic interaction. Section 3 describes the data and empirical evidence. Section 4 discusses the identification and estimation strategies. Section 5 presents the model estimates. Section 6 quantifies the importance of performance discussion to skill development. Section 7 discusses the concern of reverse causality in estimation and conducts a bounding analysis regarding the bias. Section 8 concludes.

1.2 Model

This section presents a dynamic model of parent-child strategic interaction that determines educational investments in child cognitive skill development during the first half of adolescence (ages 10 to 15).[11] The model serves for two purposes. First, it characterizes how parents strategically impact child learning behavior through performance discussion. Parents and their child play a Stackelberg game by moving sequentially. Second, it investigates under what conditions parents choose to discuss performance with their child and distinguishes various determinants of differential discussion decisions across population.

The model characterizes parent-child interaction in skill development with three features. First, children are active decision-makers who determine their own effort in learning. Second, parents can influence skill formation through two channels: one is a standard direct channel where parents make monetary investment in children's education;[12] the other is a novel indirect channel where parents discuss performance with their children to influ-

[11] According to the World Health Organization, adolescence is the phase of life from ages 10 to 19. The model starts in the beginning of adolescence and ends in the last year of middle school - the end of compulsory education in South Korea. The curriculum in compulsory education is standard nationwide and thus the learning process is comparable across schools. After compulsory schooling, a small portion of adolescents will not continue their education and the rest would enter different tracks of education, including academic and vocational ones. Therefore, cognitive development in the second half of adolescence is much more diverse and is beyond the scope of this paper.

[12] Parental investments in education generally consist of time and money. Existing literature has shown that 1) parental time spent with the child declines with the age of the child and parental time investment in adolescents is less than half of that in preschoolers (Del Boca et al., 2014), 2) adolescents' cognitive skills do not increase with parental time (Del Boca et al., 2017), and 3) parental monetary investment is more productive for adolescents than preschoolers (Del Boca et al., 2014). In addition, one typical form of monetary investment - private tutoring - paid by parents is widespread in South Korea. Thus, this paper focuses on parental educational expenditure instead of time investment during adolescence.

ence child effort choices.[13] Third, skill development is accumulative, and child skill is a "public" good valued by the child and her parents. Fourth, the model allows for correlated heterogeneity in unobserved preferences and abilities to account for differential interaction patterns across population.

1.2.1 Environment

In each household, parents[14] may have multiple children, but only one child per household is modeled as an active agent. Despite not explicitly modeling potential siblings' interactions with parents,[15] parent-child interaction patterns are allowed to vary with family composition[16] in a flexible way by considering rich heterogeneity in preferences. Households are heterogeneous in child learning efficiency, parental preferences, and child preferences, which are common knowledge within household but unobserved to researchers. These unobserved household traits are allowed to be correlated.

1.2.2 Timing of Actions

The model starts in the outset of adolescence. Household i is endowed with child learning ability A_i and initial child skill $\theta_{i,1}$. At the beginning of each period t, existing child skill $\theta_{i,t}$ is observed by parents and the child in household i. Then, after receiving household income $Y_{i,t}$ and observing realized shocks to consumption preference $\epsilon^{\delta}_{i,t}$ and to performance discussion cost $\epsilon^{\xi}_{i,t}$, parents i make two decisions: monetary investment in the child's education $M_{i,t}$ and whether to discuss performance with the child $S_{i,t}$.

[13] Parental performance discussion is not considered as a time investment for two reasons. First, performance discussion is not an ingredient in child learning process or a direct educational input in skill production. It affects child learning through changing child effort. Second, although communication with children takes the form of time, it is not likely to take a considerable amount of time, so time is not likely to be a major constraint on whether parents choose to do it.

[14] Father and mother are assumed to be a unitary decision-maker.

[15] Another concern about existing siblings is that siblings interact with each other, which poses a challenge to isolate parental impact on the child in focus from siblings' impact. To address this issue, I restrict the sample to only-child households as a robustness check in the appendix.

[16] Family composition may affect how parents interact with their child. First, trade-offs between quantity and quality of children may lead to different interaction patterns by the number of siblings. Second, parenting strategy may vary by birth order if parents want to set up a good example for younger children by being stricter to the first-born.

Given parents' choices of $M_{i,t}$ and $S_{i,t}$ and realized preference shock to leisure $\epsilon^{\lambda}_{i,t}$, child i chooses her own effort $L_{i,t}$. After investment decisions are made, production shock $\eta_{i,t}$ is realized. At the end of period t, the new child skill $\theta_{i,t+1}$ is produced with current skill and investments from parents and the child; the production process is affected by child learning ability and temporary production shock. The development process lasts for T periods until the end of compulsory education. The final product is the terminal child skill $\theta_{i,T+1}$.

1.2.3 Skill Technology

Child cognitive skill evolves according to the skill technology (1.1). The new cognitive skill $\theta_{i,t+1}$ is produced with educational inputs including current cognitive skill $\theta_{i,t}$, child effort $L_{i,t}$, and parental investment in education $M_{i,t}$. The skill technology is individual-specific to the extent that total factor productivity (TFP) A_i is heterogeneous across population. Assuming individual-specific TFPs is a flexible way of capturing all time-invariant observed and unobserved heterogeneity in child and family characteristics[17] and omitted inputs. Skill production is also stochastic given the random production shock $\exp(\eta_{i,t})$, i.i.d. across individuals and periods.

$$\theta_{i,t+1} = A_i \theta_{i,t}^{\alpha_1} (L_{i,t})^{\alpha_2} (M_{i,t})^{\alpha_3} \exp(\eta_{i,t}) \tag{1.1}$$

The form of Cobb-Douglas technology captures two features: diminishing returns of inputs and positive complementarity among inputs. The idea of diminishing returns reflects that child learning capacity is limited and when approaching the limit, the marginal return of learning is minimal. The feature of positive complementarity is twofold. First, children with higher learning efficiency (captured by TFP A_i) make better use of educational inputs. Second, educational inputs reinforce each other.

All inputs in skill technology (1.1) take only positive values and equation (1.2) is the

[17] For example, observed demographics commonly included as controls in production function (child gender, the number of siblings, parental education level, etc.) and unobserved traits like child ability to learn.

logarithm transformation of the technology.

$$\ln \theta_{i,t+1} = \ln A_i + \alpha_1 \ln \theta_{i,t} + \alpha_2 \ln L_{i,t} + \alpha_3 \ln M_{i,t} + \eta_{i,t} \tag{1.2}$$

The linear formation simplifies identification and estimation of parameters. Two things worth noting. First, individual-specific (log) TFP $\ln A_i$ is unobserved and is allowed to correlate with initial skill $\theta_{i,1}$ and household preferences. Therefore, all educational inputs, including current skill $\theta_{i,t}$ and investments from parents $M_{i,t}$ and from the child $L_{i,t}$, are endogenous to $\ln A_i$. Second, in contrast, the mean-zero production shock $\eta_{i,t}$ is assumed to realize after investments are made, thus uncorrelated with all other inputs $\{\theta_{i,t}, L_{i,t}, M_{i,t}\}$.

1.2.4 Flow Utility

In each period t, parents i receive flow utility from current child skill $\theta_{i,t}$, consumption $(Y_{i,t} - M_{i,t})$, and performance discussion $S_{i,t}$. Consumption refers to household income net of monetary investment in child education. Performance discussion is a binary choice, $S_{i,t} \in \{0, 1\}$.

$$U^p_{i,t}(\theta_{i,t}, Y_{i,t}, M_{i,t}, S_{i,t}) = \ln \theta_{i,t} + \delta_{i,t} \ln(Y_{i,t} - M_{i,t}) - \xi_{i,t} S_{i,t} \tag{1.3}$$

Parental preference for child skill is normalized to be 1. Parents have individual-specific preference for consumption (relative to skill) δ_i and experiences random shocks to consumption in each period $\epsilon^{\delta}_{i,t}$.[18]

$$\delta_{i,t} = \delta_i \epsilon^{\delta}_{i,t}$$

This paper takes a simplified way of modeling performance discussion cost by summarizing it in one parameter ξ. It includes direct cost associated with performance discussion, such as attention and time spent, and indirect cost resulting from altruism to the child as

[18] $\epsilon^{\delta}_{i,t}$ follows a log-normal distribution with mean of 0 and standard deviation of σ_δ and is i.i.d. across individuals and periods. Equivalently, $\delta_{i,t}$ follows a log-normal distribution with mean of $\ln \delta_i$ and standard deviation of σ_δ.

performance discussion lowers child utility. Parents have individual-specific performance discussion cost (relative to skill) ξ_i and experiences random shocks to discussion cost in each period $\epsilon^{\xi}_{i,t}$.[19]

$$\xi_{i,t} = \xi_i \epsilon^{\xi}_{i,t}$$

Heterogeneity in performance discussion costs consists of two parts: one is the permanent difference in mean cost that captures persistent variation in performance discussion tendency across population; the other is the transitory shock to cost that captures the within-individual variation in performance discussion choice over time. The average cost ξ_i is parameterized to be correlated with parental preference for consumption (δ_i) and other observed household characteristics, including parental education (x_{1i}), birth order (x_{2i}), and child gender (x_{3i}), in equation 1.4. The correlation of performance discussion cost and consumption preference reflects the correlation of unobserved heterogeneity in parents' traits and allows for correlation between parents' two choices.[20] The systematic difference in performance discussion cost by parental education suggests unequal communication skills of parents by education. Performance discussion cost may vary by birth order if parents have incentives to set up a good example for younger children by being stricter to the first-born.[21] Gender difference in performance discussion cost allows for suitable parenting strategies in line with the gender of child.

$$\xi_i = \pi_0 + \pi_1 \delta_i + \pi_2 \mathbf{1}\{x_{1i} \geq \text{college}\} + \pi_3 \mathbf{1}\{x_{2i} = \text{first child}\} + \pi_4 \mathbf{1}\{x_{3i} = \text{girl}\} \tag{1.4}$$

[19] $\epsilon^{\xi}_{i,t}$ follows a log-normal distribution with mean of 0 and standard deviation of σ_ξ and is i.i.d. across individuals and periods. Equivalently, $\xi_{i,t}$ follows a log-normal distribution with mean of $\ln \xi_i$ and standard deviation of σ_ξ.

[20] While it is restrictive to assume perfect correlation of performance discussion cost ξ_i and preference for consumption δ_i, it is hard to allow for imperfect correlation because a more flexible joint distribution is not well-identified from data given the binary observation of performance discussion. Besides, it does not impose strong assumption of perfect correlation between the observed tendency of investment (fraction of income spending on education) and probability of performance discussion, because the perfect correlation of performance discussion cost ξ_i and consumption preference δ_i does not result in perfect correlation of realized performance discussion cost $\xi_{i,t}$ and consumption preference $\delta_{i,t}$ due to transitory preference shocks.

[21] For example, Hotz and Pantano (2015) find that parents tend to impose more stringent disciplinary rules on their earlier-born children in order to deter bad behaviors of later-born offspring.

Child i receives flow utility from current skill $\theta_{i,t}$ and the interaction of leisure $(\bar{L}-L_{i,t})$ and performance discussion $S_{i,t}$. Leisure is defined as the idle time in class, which is total class time $\bar{L}$ net of class effort $L_{i,t}$ measured by time paying attention.

$$U^c_{i,t}(\theta_{i,t}, L_{i,t}, S_{i,t}) = \ln\theta_{i,t} + \lambda_{i,t}(1-\tau S_{i,t})\ln(\bar{L}-L_{i,t}) \tag{1.5}$$

Child preference for skill is normalized to be 1. Children have individual-specific preference for leisure (relative to skill) λ_i and experiences random shocks to leisure in each period $\epsilon^{\lambda}_{i,t}$.[22]

$$\lambda_{i,t} = \lambda_i \epsilon^{\lambda}_{i,t}$$

Performance discussion with parents decreases the child's value from leisure if $\tau > 0$. It can be interpreted as parents convey the importance of good performance to their child through such discussion and foster internalized motivation for effort.[23]

The development process lasts for T periods and the final product is the terminal child skill $\theta_{i,T+1}$. Parents and the child receive terminal values from final child skill:

$$\psi^p \ln\theta_{i,T+1} \text{ for parents}$$
$$\psi^c \ln\theta_{i,T+1} \text{ for child}$$

1.2.5 Dynamic Problems

This section formulates the dynamic problem faced by parents and the child. As the timing of actions suggests, parents and the child play a Stackelberg game in each period, where parents are the leader and the child follows. When analyzing a Stackelberg game, it is more straightforward to solve backwards by starting with the follower's problem.

[22] $\epsilon^{\lambda}_{i,t}$ follows a log-normal distribution with mean of 0 and standard deviation of σ_λ and is i.i.d. across individuals and periods. Equivalently, $\lambda_{i,t}$ follows a log-normal distribution with mean of $\ln\lambda_i$ and standard deviation of σ_λ.

[23] Decreasing the value of leisure is decreasing the cost of effort.

Child's Problem

In each period t, child i observes existing skill $\theta_{i,t}$ and takes as given parents' choices of educational investment and performance discussion $\{M_{i,t}, S_{i,t}\}$. Then, given realized preference shock to leisure, she chooses class effort $L_{i,t}$ to maximize her value function subject to the skill technology.

$$\begin{aligned} V^c_{i,t}(\theta_{i,t}, \lambda_{i,t} | M_{i,t}, S_{i,t}) = \max_{L_{i,t}|M_{i,t},S_{i,t}} & \ln \theta_{i,t} + \lambda_{i,t}(1 - \tau S_{i,t}) \ln(\bar{L} - L_{i,t}) \\ & + \beta^c \mathbb{E} V^c_{i,t+1}(\theta_{i,t+1}, \lambda_{i,t+1} | M_{i,t}, S_{i,t}) \\ s.t. \quad & \ln \theta_{i,t+1} = \ln A_i + \alpha_1 \ln \theta_{i,t} + \alpha_2 \ln L_{i,t} + \alpha_3 \ln M_{i,t} + \eta_{i,t} \end{aligned}$$

where β^c is the discount factor; the expectation operator $\mathbb{E}$ is over the mean-zero production shock $\eta_{i,t}$.

Solutions. The Child's problem is solved backwards and solution details are described in appendix A.1.1. Optimal child effort $L^*_{i,t}$ depends on parental performance discussion $S_{i,t}$, child preference parameters, and production parameters, but not on state variable $\theta_{i,t}$ or parental educational investment $M_{i,t}$.[24] The policy function (1.6) indicates two sources of variation in child effort. First, heterogeneous preferences for leisure $\lambda_{i,t}$ lead to cross-sectional and within-child variation in effort. Leisure preference can be interpreted as an inverse measure of diligence: a lower preference for leisure means child being more diligent, thus making higher effort (conditional on performance discussion decision). In addition, parental performance discussion increases optimal child effort through decreasing the value of child leisure, which adds to cross-sectional and within-child variation in effort.

$$L^*_{i,t}(S_{i,t}) = \frac{\bar{L} \beta^c \Gamma^c_{t+1} \alpha_2}{\lambda_{i,t}(1 - \tau S_{i,t}) + \beta^c \Gamma^c_{t+1} \alpha_2} \tag{1.6}$$

where the sequence of $\{\Gamma^c_t\}_{t=1}^{T+1}$ is defined recursively and common across individuals. Γ^c_t

[24] The independence among educational inputs $\{\theta_{i,t}, L_{i,t}, M_{i,t}\}$ is a result of additive separability of terms in utility and production function. See proofs in appendix A.1.1.

is interpreted as the child's marginal utility from current (log) skill in period t.

$$\begin{aligned}
\Gamma^c_{T+1} &= \psi^c \\
&... \\
\Gamma^c_t &= 1 + \beta^c \Gamma^c_{t+1} \alpha_1 \\
&... \\
\Gamma^c_1 &= 1 + \beta^c \Gamma^c_2 \alpha_1
\end{aligned}$$

Parents' Problem

In each period t, parents i observe the child's existing skill $\theta_{i,t}$ and receive household income $Y_{i,t}$. Then, given realized shocks to consumption preference and to performance discussion cost and the knowledge of the child's response function (1.6) (up to λ_i),[25] they make decisions on monetary investment in child education $M_{i,t}$ and on whether to discuss performance with the child $S_{i,t}$.

$$V^p_{i,t}(\theta_{i,t}) = \max_{M_{i,t}, S_{i,t} | L^*_{i,t}(S_{i,t})} \ln \theta_{i,t} + \delta_{i,t} \ln(Y_{i,t} - M_{i,t}) - \xi_{i,t} S_{i,t} + \beta^p \mathbb{E} V^p_{i,t+1}(\theta_{i,t+1})$$

$$s.t. \quad \ln \theta_{i,t+1} = \ln A_i + \alpha_1 \ln \theta_{i,t} + \alpha_2 \ln L_{i,t} + \alpha_3 \ln M_{i,t} + \eta_{i,t}$$

where β^p is the discount factor; the expectation operator $\mathbb{E}$ is over the mean-zero production shock $\eta_{i,t}$.

Solutions. Parents' problem is solved backwards and solution details are described in appendix A.1.2. Optimal parental educational investment $M^*_{i,t}$ is a function of household income, parental preference parameters, and production parameters, but it is independent of state variable $\theta_{i,t}$ and performance discussion decision $S_{i,t}$.[26] According to the policy function (1.7), variation in parental investment comes from two sources. First, richer parents can afford more investment. Second, heterogeneous preference for consumption

[25] Parents make their decisions without observing preference shocks to child leisure.

[26] The independence is a result of additive separability of terms in utility and production function. See proofs in appendix A.1.2.

$\delta_{i,t}$ is an inverse measure of investment tendency in the child's education. Parents placing a low value on consumption tend to invest more in education for their child conditional on family income.

$$M^*_{i,t} = \frac{\beta^p \Gamma^p_{t+1} \alpha_3}{\delta_{i,t} + \beta^p \Gamma^p_{t+1} \alpha_3} Y_{i,t} \tag{1.7}$$

where the sequence of $\{\Gamma^p_t\}_{t=1}^{T+1}$ is defined recursively and common across individuals. Γ^p_t is interpreted as parents' marginal utility from current (log) child skill in period t.

$$\begin{aligned}
\Gamma^p_{T+1} &= \psi^p \\
\Gamma^p_T &= 1 + \beta^p \Gamma^p_{T+1} \alpha_1 \\
&\dots \\
\Gamma^p_t &= 1 + \beta^p \Gamma^p_{t+1} \alpha_1 \\
&\dots \\
\Gamma^p_1 &= 1 + \beta^p \Gamma^p_2 \alpha_1
\end{aligned}$$

Parent's decision of performance discussion hinges on the trade-off between the performance discussion cost in current period and the benefit from a higher future skill. Optimal discussion decision $S^*_{i,t}$ depends on discussion cost, preference parameters of parents and the child, and production parameters, but not on state variable $\theta_{i,t}$ or parental educational investment $M_{i,t}$.[27] It is optimal for parents to discuss performance with the child if performance discussion results in higher utility than no performance discussion, i.e., a greater conditional value with discussion $V^p_{i,t}(\theta_{i,t}|S_{i,t} = 1)$ than without discussion

[27] The independence of choices and state variable is a result of additive separability of terms in utility and production function. See proofs in appendix A.1.2.

$V^p_{i,t}(\theta_{i,t}|S_{i,t}=0)$.

$$S^*_{i,t} = \mathbf{1}\{V^p_{i,t}(\theta_{i,t}|S_{i,t}=1) > V^p_{i,t}(\theta_{i,t}|S_{i,t}=0)\}$$
$$= \mathbf{1}\{\underbrace{-\xi_{i,t}}_{\text{discussion cost}} + \underbrace{\beta^p\Gamma^p_{t+1}\alpha_2 \ln \frac{\lambda_i + \beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_i(1-\tau) + \beta^c\Gamma^c_{t+1}\alpha_2}}_{\text{discussion benefit from higher future skill}} > 0\}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function; $\xi_{i,t}$ is the cost of performance discussion, which follows a log-normal distribution with mean of $\ln \xi_i$ and standard deviation of σ_ξ.

$$F_\xi = \Phi\left(\frac{\ln \xi_{i,t} - \ln \xi_i}{\sigma_\xi}\right)$$

The second term is the benefit of performance discussion, denoted as $B_{i,t}$.

$$B_{i,t} = \beta^p\Gamma^p_{t+1}\alpha_2 \ln \frac{\lambda_i + \beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_i(1-\tau) + \beta^c\Gamma^c_{t+1}\alpha_2}$$

Let $q_{i,t}$ be the probability of parents i choosing to discuss performance with the child in period t when the benefit outweighs the cost: $q_{i,t} = \mathbb{P}(S^*_{i,t} = 1) = \mathbb{P}(\xi_{i,t} < B_{i,t})$. Given the distribution of discussion cost, the probability $q_{i,t}$ is calculated below.

$$q_{i,t} = \Phi\left(\frac{\ln B_{i,t} - \ln \xi_i}{\sigma_\xi}\right) \tag{1.8}$$

The probability of performance discussion is heterogeneous across households for two reasons. First, the benefit from performance discussion $B_{i,t}$ varies by the child's preference for leisure. Moreover, performance discussion cost $\xi_{i,t}$ varies across population.

1.3 Empirical Evidence

1.3.1 Sample Construction

I utilize data from the Korean Education Longitudinal Study (KELS-13) that traces the cognitive development of a nationally representative cohort of children in South Korea from primary to post-secondary education.[28] KELS-13 is a comprehensive annual survey that provides uniquely detailed information on parent-child interactions, child learning behaviors and progress in achievement tests, and educational resources from parents and schools. It has a two-stage stratified sampling design in which (1) elementary schools are randomly sampled within each of four types of administrative units[29] proportionally to the population size, and then (2) approximately 50 students (all if less than 50) in the fifth grade are randomly sampled within the selected schools. In the first wave of 2013, 7,287 fifth graders (along with their parents and teachers) in 242 schools across the nation respond to the baseline survey. 90 percent (6,517) of the original respondents are successfully tracked in the subsequent four waves until the end of compulsory schooling.[30] Separate surveys are administered to students, parents, teachers, and school principals to collect a wide range of family, class, and school information in each wave. KELS also administers achievement tests in Math, English, and Korean to students in the first five waves.[31]

The sample used in estimation consists of 2,689 public school students,[32] for whom I observe demographic information,[33] family financial resources, educational inputs, parent-child interactions, and test scores in three core subjects in the first five waves. The following table summarizes the main variables used in the estimation.

[28]KELS-13 is conducted by the Korean Educational Development Institute, a government-funded educational research agency.

[29]Administrative units are classified and ordered by the level of urbanization: the capital city (Seoul), metropolitan cities, small and medium-sized cities, and towns.

[30]Compulsory schooling in South Korea covers six-year primary school and 3-year middle school. The curriculum of compulsory education follows the same standard nationwide.

[31]Math, English, and Korean are the core subjects for the national college entrance examination and are considered as good measures of cognitive skills.

[32]The estimation sample accounts for 41 percent of the 6,517 respondents who are successfully traced for the first five waves or 37 percent of the 7,287 original respondents.

[33]Demographic characteristics include age and gender of student, number of siblings, and age and education level of both parents.

Variable	*Description*	*Survey Years*
$z_{i,t}^{j}$	Child test scores in subject j: $j \in \{\text{Math, English, Korean}\}$	2013-2017
$L_{i,t}$	Child effort: percentage of time paying attention in class	2014-2017
$S_{i,t}$	Performance discussion: Parent-child discussion about school performance	2014-2017
$M_{i,t}$	Parental educational investment: annual educational expenditure	2014-2017
$Y_{i,t}$	Household income: annual total family income	2014-2017

Children's cognitive skills are latent and measured imperfectly with test scores in Math, English, and Korean.[34] The achievement tests are designed to track children's learning progress and test scores are directly comparable across grades. Measurement errors in test scores are corrected by combining information from all available measures, which will be described in details in Section 1.4.1.

I construct a continuous measure of child effort on a scale from 0 to 100 percent. In each survey wave, children self-report the time that they concentrate and pay attention in a typical Math/English/Korean class in the past year. I calculate the average concentrating time in class of the three subjects and divide by the length of a typical class.[35] I define the percentage of concentrating time in a typical class as the measure of child effort.

I construct a binary variable of performance discussion from the reported parent-child interaction. Children are asked how often their parents talk to them about their performance in school and encourage them to perform well. I assign the value of 1 to

[34] Scores in each subject are standardized such that the distribution of fifth graders' scores has mean 0 and standard deviation 1.

[35] In South Korea, classes in elementary schools last for 40 minutes and classes in middle schools last for 50 minutes.

performance discussion for a child if she reports that her parents are likely or very likely to ask about her school performance in the past year. By regularly discussing about school performance, parents convey the value of schooling to the child and show their interest in the child's performance, which influences the child's effort choice.[36]

Parents report their average monthly expenditure on the surveyed child's education for the past year in each wave. Education expenditure is defined as household spending on all education-related activities, including academic-oriented private tutoring, non-academic extracurricular programs, educational supplies, etc. I keep only children enrolled in public schools in all survey years, for whom no tuition is charged by schools.[37] I convert the monthly educational expenditure to annual expenses to match the model period. In addition, parents report their monthly household income in a typical month for the past year in each wave, which is converted to annual income.

1.3.2 Summary Statistics

Table 1.1 reports descriptive statistics for the sample used to estimate the model. Panel (A) presents demographic information. It is a gender-balanced sample with 52 percent of girls. In the initial wave of 2013, all students are in fifth grade; the average ages of their parents are 40.60 for mothers and 43.25 for fathers, respectively. Fathers are more educated than mothers on average. 57 percent of mothers and 65 percent of fathers have received college education.

Panel (B) summarizes family financial resources, educational investments, and parent-child interaction in waves of 2014-2017 (corresponding to the model's initial to final periods). The average annual household income is \$49,974.32.[38] Parents spend an average of \$4,609.22 or 10.15 percent of total income on the surveyed child's education every year.[39]

[36] Although parents cannot directly decide learning behaviors for their child, they can influence the child's choices. In the surveyed sample, almost all children agree (at least to some extent) that they study hard because they do not want to disappoint their parents.

[37] Public schooling during compulsory education is subsidized by government and free to all children.

[38] Household income and educational expenditure are converted to 2010 US dollars.

[39] The average ratio of educational expenditure over income is larger than the ratio of average expenditure over average income, because the expenditure-income ratio for the highest quartile of income group is lower than that for the lower quartiles.

Table 1.1: Descriptive Statistics for Estimation Sample

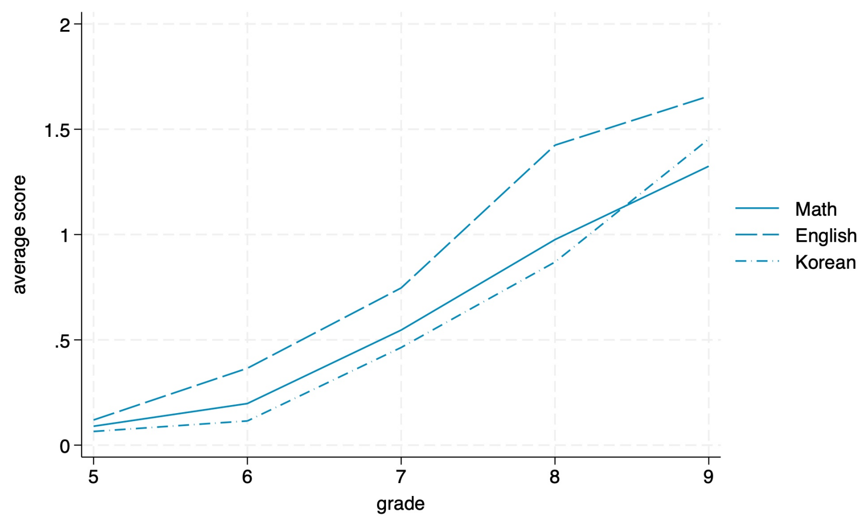
	Mean	Std. Dev.	No. of Obs.
(A) Demographic Information in 2013 KELS			
Fraction of female	0.52		2,689
Fraction of first-born child	0.47		2,661
No. of Siblings	1.18	(0.67)	2,661
Mother's age	40.60	(3.53)	2,544
Father's age	43.25	(3.76)	2,279
Fraction of College-Educated Mothers	0.57		2,634
Fraction of College-Educated Fathers	0.65		2,530
(B) Main Variables in 2014-2017 KELS			
Annual household income (in 2010 USD)	49,974.32	(34,636.81)	10,756
Grade 6 in 2014	49,527.06	(30,993.40)	2,689
Grade 7 in 2015	48,775.35	(30,352.31)	2,689
Grade 8 in 2016	50,286.94	(36,641.84)	2,689
Grade 9 in 2017	51,307.95	(39,650.73)	2,689
Annual educational expense (in 2010 USD)	4,609.22	(2,831.14)	10,756
Grade 6 in 2014	4,079.52	(2,308.01)	2,689
Grade 7 in 2015	4,569.01	(2,639.08)	2,689
Grade 8 in 2016	4,819.70	(3,306.00)	2,689
Grade 9 in 2017	4,968.64	(2,897.09)	2,689
% of income spending on education	10.15	(5.42)	10,756
Grade 6 in 2014	9.16	(5.08)	2,689
Grade 7 in 2015	10.24	(5.37)	2,689
Grade 8 in 2016	10.48	(5.46)	2,689
Grade 9 in 2017	10.73	(5.61)	2,689
% of time paying attention in class	69.37	(18.65)	10,756
Grade 6 in 2014	69.89	(16.41)	2,689
Grade 7 in 2015	64.92	(17.94)	2,689
Grade 8 in 2016	63.80	(18.72)	2,689
Grade 9 in 2017	62.37	(19.62)	2,689
Fraction of discussing performance	0.55		10,756
Grade 6 in 2014	0.54		2,689
Grade 7 in 2015	0.56		2,689
Grade 8 in 2016	0.57		2,689
Grade 9 in 2017	0.53		2,689

Source: KELS-13
Notes: Means and standard deviations are computed with valid observations.

The average child effort measured by the percentage of time paying attention in class is 69.37 percent. About 55 percent of parents regularly discuss performance with their child.

Figure 1.1 plots the average test scores in each wave. It is not surprising that the average scores are rising over time, because the tests are designed to track students' learning progress and are supposed to depict the developmental process of cognitive skill. From fifth grade to ninth grade, the average scores increase by 1.23-1.53 standard deviation.

Figure 1.1: Time Trend in Test Scores



Notes: This figure shows the average scores in each subject (Source: KELS). The x-axis refers to the grade of children.

1.3.3 Performance Discussion and Child Effort

Table 1.2 examines the relationship of performance discussion and class effort. Column (1) reports a simple OLS regression of time paying attention in class on performance discussion (controlling for grade fixed effects).[40] Without additional controls, performance discussion is associated with 18 p.p. increase in concentrating time in class. When adding demographic variables as controls in column (2), the coefficient of performance discussion on class effort drops to 16.52 p.p., suggesting that some household characteristics (e.g., parental education) are positively correlated with both performance discussion and class effort. When replacing observed demographic variables with individual fixed effects in column (3), the coefficient of performance discussion on class effort drops further to 12.78

[40] Grade fixed effects account for the observed time trend in class effort.

p.p., suggesting that individual heterogeneity that causes selection bias is not fully accounted for by observed characteristics. Exploiting within-individual variation, I find a significantly positive effect of performance discussion on class effort.

Table 1.2: Regression of Child Effort on Performance Discussion

Outcome	% time paying attention in class		
	(1)	(2)	(3)
Performance Discussion	18.00***	16.52***	12.78***
	(1.66)	(1.70)	(4.80)
No discussion mean	63.66	63.66	63.66
Grade FE	Yes	Yes	Yes
Demographics		Yes	
Individual FE			Yes
Observations	13,303	13,303	13,303
Individuals	2,689	2,689	2,689

Standard errors clustered at individual level in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Source: KELS-13
Notes: The table reports the regression results of class effort, measured by percent of time paying attention in class, on performance discussion. Demographic variables include child gender, birth order, and the education levels of mothers and fathers.

1.3.4 Educational Inequality

The model allows for rich heterogeneity among households to capture differential household behaviors and diverse skill development paths. One policy-relevant aspect the model speaks to is the educational inequality in both investments and outcomes across population. Panel (B) in table 1.3 provides empirical evidence for gaps in educational inputs between households in high- and low-socioeconomic groups.[41]

[41] I measure socioeconomic status (SES) by maternal education and define households with a college-educated mother as high SES.

Table 1.3: Educational Inequality

	Low-SES		High-SES		
	Mean	SD	Mean	SD	Difference
(A) Demographic Information in 2013 KELS					
Fraction of female	0.52		0.52		0.00
Fraction of first-born child	0.41		0.52		-0.11***
No. of Siblings	1.26	(0.77)	1.12	(0.58)	0.15***
Mother's age	40.50	(3.85)	40.66	(3.15)	-0.16*
Father's age	43.27	(4.22)	43.21	(3.35)	0.06
(B) Main Variables in 2014-2017 KELS					
Annual household income (in 2010 USD)	42,676.50	(26,695.49)	55,747.59	(37,396.54)	-13,071.10***
Annual educational expenditure (in 2010 USD)	3,931.25	(2,454.70)	5,176.42	(2,995.08)	-1,245.17***
Percentage of income spending on education (%)	10.21	(5.52)	10.13	(5.34)	0.09
Percentage of time paying attention in class (%)	62.94	(18.65)	67.20	(17.94)	-4.25***
Fraction of parents discussing performance with child	0.49		0.60		-0.10***

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

Source: KELS-13

Notes: The table compares households from low- and high-socioeconomic groups in terms of demographics, financial resources, educational inputs, and parent-child interaction. Socioeconomic Status (SES) is measured by maternal education: a household with a college-educated mother is defined as high SES.

High-SES households have 30.6 percent more annual income than that of the low-SES. Therefore, the high-SES group can afford more educational investment: their average spending on children's education is 31.7 percent more than that of the low-SES. Besides monetary investment, high-SES parents are also more involved in their children's schooling: they are 10 percentage points more likely to have performance discussion with their children than the low-SES. The SES gradient is also salient in children's behaviors: high-SES children have 4.35 percentage points more concentrating time in class than the low-SES.

Figure 1.2 depicts the persistent and widening gap in test scores between children from high- and low-socioeconomic groups. High-SES children not only start with higher test scores in the initial year, but also improve their achievement at higher rates. As a result, the math score gap of children in high- and low-SES families increases from 0.38 to 0.54 standard deviation in four years.

1.4 Estimation

1.4.1 Measurement Parameters for Latent Skill

This section describes the identification of the measurement model of child skill. Cognitive skill $\theta_{i,t}$ is a latent variable and is measured with errors. J_θ measures are available: $z^j_{i,t}$ is the j^{th} measure of child i's latent skill in period t. Measurement model is log-linear in latent skill.[42]

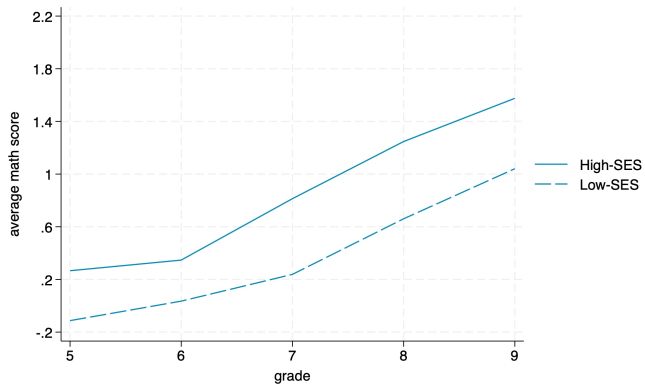
$$z^j_{i,t} = \mu^j_t + \omega^j_t \ln \theta_{i,t} + \upsilon^j_{i,t} \quad \text{for } t = 1, 2, .., T \text{ and } j = 1, 2, ..., J_\theta \tag{1.9}$$

where μ^j_t and ω^j_t represent the location and scale of the j^{th} measure in period t, respectively; $\upsilon^j_{i,t}$ is the unobserved measurement error, for which $\mathbb{E}(\upsilon^j_{i,t}) = 0$ without loss of generality.

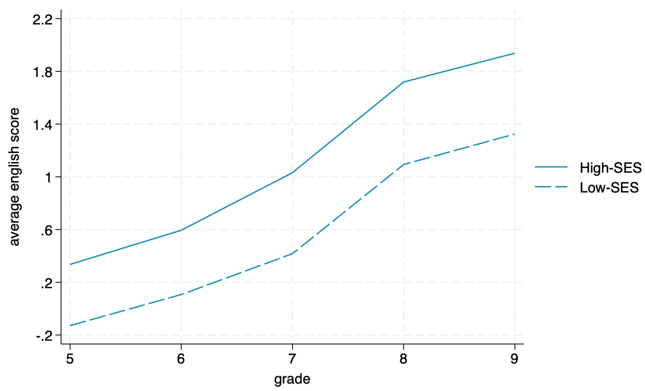
The measurement parameters for all measures in all periods are identified when three conditions are satisfied: 1) some independence restrictions on the measurement errors,

[42] Latent cognitive skill takes non-negative values, while measures of skill can be any real numbers.

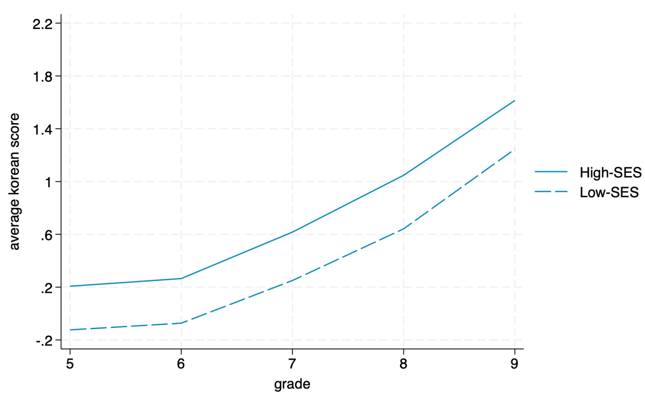
Figure 1.2: Score Gap between High- and Low-SES



(a) Math



(b) English



(c) Korean

Notes: This figure shows the score gap between high- and low-SES students in each subject (Source: KELS). The x-axis refers to the grade of children.

2) at least one *age-invariant* measure, and 3) some normalization of initial skill. First, measurement errors are assumed to be independent of the latent variable and of each other. Assumption 1(a) is that for all individuals and all measures, measurement errors are independent of latent variables in the same and different periods. Assumption 1(b) is that for all individuals, measurement errors are independent across measures in the same and different periods.

Assumption 1. *Independence assumptions of measurement errors:*

(a) $\upsilon_{i,t}^{j} \perp\!\!\!\perp \theta_{i,t'}$ *for all* i*, all* j*, all* t*, and all* t'

(b) $\upsilon_{i,t}^{j} \perp\!\!\!\perp \upsilon_{i,t'}^{j'}$ *for all* i*, all* $j \neq j'$*, all* t*, and all* t'

Second, the idea of *age-invariant* measures follows the definition in Agostinelli and Wiswall (2025). The intuition is that two children of different ages with the same level of latent skill would on average perform equally well in an age-invariant measure. Definition 1 formalizes the argument.

Definition 1. *A pair of measures* $z_{i,t}^{j}$ *and* $z_{i,t'}^{j}$ *is age-invariant if* $\mathbb{E}(z_{i,t}^{j}|\theta_{i,t} = p) = \mathbb{E}(z_{i,t'}^{j}|\theta_{i,t'} = p)$ *for all* $t \neq t'$ *and all* $p \in \mathbb{R}_{++}$*.*

Definition 1 along with Assumption 1(a) establishes the key feature of an age-invariant measure z^j that its measurement parameters are constant in all periods: $\mu_t^j = \mu^j$ and $\omega_t^j = \omega^j$. Without loss of generality, let the age-invariant measure be the first measure so that $\mu_t^1 = \mu^1$ and $\omega_t^1 = \omega^1$.

Lastly, the following normalization of initial skill $\theta_{i,1}$ is necessary, because skill does not have a natural location or scale. Beyond this, no further restrictions are imposed on latent skill in latter periods ($t > 1$).[43]

Normalization. *For initial skill* $\theta_{i,1}$*,*

(a) $\mathbb{E}(\ln \theta_{i,1}) = 0$*, the mean logarithm of initial skill is 0.*

[43] Biases could arise if latent variables are re-normalized in all periods of a dynamic model (Agostinelli and Wiswall, 2016).

(b) $\mathbb{V}(\ln\theta_{i,1}) = 1$*, the variance of logarithm of initial skill is 1.*

Under the above three conditions, I describe the identification of the scale and location parameters for all measures in all periods. In initial period ($t = 1$), measurement parameters are identified given the initial normalization and independence assumptions of measurement errors. The location parameters μ_1^j for all measures are identified given the initial normalization of $\mathbb{E}(\ln\theta_{i,1}) = 0$ and mean-zero measurement errors.

$$\mu_1^j = \mathbb{E}(z_1^j) - \omega_1^j \mathbb{E}(\ln\theta_{i,1}) - \mathbb{E}(\upsilon_{i,1}^j) = \mathbb{E}(z_1^j)$$

The covariance of different measures identifies the scale parameters ω_1^j for all measures.

$$\begin{aligned} \omega_1^j &= \sqrt{\frac{Cov(z_{i,1}^j, z_{i,1}^{j'})Cov(z_{i,1}^j, z_{i,1}^{j''})}{Cov(z_{i,1}^{j''}, z_{i,1}^{j'})}} \\ \text{as} \quad Cov(z_{i,1}^j, z_{i,1}^{j'}) &= \omega_1^j \omega_1^{j'} \mathbb{V}(\ln\theta_{i,1}) = \omega_1^j \omega_1^{j'} \\ Cov(z_{i,1}^j, z_{i,1}^{j''}) &= \omega_1^j \omega_1^{j''} \mathbb{V}(\ln\theta_{i,1}) = \omega_1^j \omega_1^{j''} \\ Cov(z_{i,1}^{j''}, z_{i,1}^{j'}) &= \omega_1^{j''} \omega_1^{j'} \mathbb{V}(\ln\theta_{i,1}) = \omega_1^{j''} \omega_1^{j'} \end{aligned}$$

where $z_{i,1}^j$, $z_{i,1}^{j'}$, and $z_{i,1}^{j''}$ are three different measures for initial skill $\theta_{i,1}$; their covariance is function of scale parameters given the initial normalization of $\mathbb{V}(\ln\theta_{i,1}) = 1$ and independence assumptions of measurement errors.

The first measure z^1 is age-invariant so that the location μ^1 and scale ω^1 recovered in initial period remain the same in later periods ($t > 1$). The measurement parameters for all other measures in later periods are identified given the age-invariant measure and independence assumptions of measurement errors. The scale parameters ω_t^j for all $j > 1$

is identified by the covariance of different measures.

$$
\begin{aligned}
\omega_t^j &= \frac{\omega^1 Cov(z_{i,t}^j, z_{i,t}^{j'})}{Cov(z_{i,t}^1, z_{i,t}^{j'})} \\
\text{as} \quad Cov(z_{i,t}^j, z_{i,t}^{j'}) &= \omega_t^j \omega_t^{j'} \mathbb{V}(\ln \theta_{i,t}) \\
Cov(z_{i,t}^1, z_{i,t}^{j'}) &= \omega^1 \omega_t^{j'} \mathbb{V}(\ln \theta_{i,t})
\end{aligned}
$$

where $z_{i,t}^j$ and $z_{i,t}^{j'}$ are two different measures other than $z_{i,t}^1$; their covariance is function of scale parameters and variance of (log) skill given independence assumptions of measurement errors.

Once the scale parameters are recovered, the location parameters μ_t^j for all $j > 1$ are identified given mean-zero measurement errors and the mean of (log) skill estimated by transforming the mean of the age-invariant measure.

$$
\begin{aligned}
\mu_t^j &= \mathbb{E}(z_{i,t}^j) - \omega_t^j \mathbb{E}(\ln \theta_{i,t}) \\
\text{as} \quad \mathbb{E}(\ln \theta_{i,t}) &= \frac{\mathbb{E}(z_{i,t}^1) - \mu^1}{\omega^1}
\end{aligned}
$$

After recovering all measurement parameters, the measures can be re-scaled as a combination of the latent skill $\ln \theta_{i,t}$ and some error by subtracting the location and dividing by the scale. $\tilde{z}_{i,t}^j$ is the re-scaled measure.

$$
\begin{aligned}
\frac{z_{i,t}^j - \mu_t^j}{\omega_t^j} &= \ln \theta_{i,t} + \frac{\upsilon_{i,t}^j - \mu_t^j}{\omega_t^j} \\
\tilde{z}_{i,t}^j &= \ln \theta_{i,t} + \tilde{\upsilon}_{i,t}^j
\end{aligned}
$$

1.4.2 Production Parameters

This section presents the identification of productivity parameters $\{\alpha_1, \alpha_2, \alpha_3\}$, permanent TFP shock $\mathbb{V}(\ln A_i)$, and temporary production shock $\mathbb{V}(\eta_{i,t})$.

Child cognitive skill $\theta_{i,t}$ is a latent variable with several error-contaminated measures. Replacing latent skill with its j^{th} re-scaled measure gives the empirical counterpart of skill technology (1.2).

$$\tilde{z}^j_{i,t+1} = \ln A_i + \alpha_1 \tilde{z}^j_{i,t} + \alpha_2 \ln L_{i,t} + \alpha_3 \ln M_{i,t} + \eta_{i,t} + \tilde{\upsilon}^j_{i,t+1} - \alpha_1 \tilde{\upsilon}^j_{i,t} \tag{1.10}$$

where $\tilde{z}^j_{i,t}$ is the re-scaled measure of skill; child effort $L_{i,t}$ and parental educational investment $M_{i,t}$ are observable; the unobserved components include TFP shock $\ln A_i$, production shock $\eta_{i,t}$, and measurement errors $\tilde{\upsilon}^j_{i,t+1}$ and $\tilde{\upsilon}^j_{i,t}$. To obtain unbiased estimates of production parameters $\{\alpha_1, \alpha_2, \alpha_3\}$, two issues need to be addressed. First, unobserved TFP shock $\ln A_i$ correlates with the measure of skill $\tilde{z}^j_{i,t}$ and investments from the child $L_{i,t}$ and parents $M_{i,t}$.[44] Second, unobserved measurement error $\tilde{\upsilon}^j_{i,t}$ correlates with the measure of skill $\tilde{z}^j_{i,t}$.[45]

To deal with the above challenges, the modified Arellano-Bond estimator that uses lags in alternate measures as instruments is employed and described below (Andrabi et al., 2011). Let $\iota^j_{i,t} = \eta_{i,t} + \tilde{\upsilon}^j_{i,t+1} - \alpha_1 \tilde{\upsilon}^j_{i,t}$. First difference equation (1.10) to remove unobserved permanent TFP shock $\ln A_i$.

$$\begin{aligned}
\tilde{z}^j_{i,t+1} - \tilde{z}^j_{i,t} &= \alpha_1(\tilde{z}^j_{i,t} - \tilde{z}^j_{i,t-1}) + \alpha_2(\ln L_{i,t} - \ln L_{i,t-1}) + \alpha_3(\ln M_{i,t} - \ln M_{i,t-1}) \\
&\quad + (\iota^j_{i,t} - \iota^j_{i,t-1}) \\
\Delta \tilde{z}^j_{i,t+1} &= \alpha_1 \Delta \tilde{z}^j_{i,t} + \alpha_2 \Delta \ln L_{i,t} + \alpha_3 \Delta \ln M_{i,t} + \Delta \iota^j_{i,t}
\end{aligned}$$

where $\Delta \iota^j_{i,t} = \iota^j_{i,t} - \iota^j_{i,t-1} = (\eta_{i,t} + \tilde{\upsilon}^j_{i,t+1} - \alpha_1 \tilde{\upsilon}^j_{i,t}) - (\eta_{i,t-1} + \tilde{\upsilon}^j_{i,t} - \alpha_1 \tilde{\upsilon}^j_{i,t-1})$.

Then, exploit the following moment conditions (1.11) to instrument for $\Delta \tilde{z}^j_{i,t}$ using one- or more-period lags of alternate measures $\tilde{z}^{j'}$.[46] The validity of alternate lagged measures $\tilde{z}^{j'}$ as instruments relies on that measurement errors are independent across measures (see

[44] The correlation of TFP and investments comes from the correlation of TFP and household preferences that determine investment choices.

[45] As assumed in the model, production shock $\eta_{i,t}$ is realized after investments are made, thus uncorrelated with contemporary educational inputs. In addition, it is reasonable to assume that production shocks are independent of measurement errors of skills.

[46] No instrument is needed for $\Delta \ln L_{i,t}$ or $\Delta \ln M_{i,t}$, because the model predicts that investment choices do not depend on current skill so that they do not correlate with previous production shocks $\eta_{i,t-1}$.

detailed proofs of the moment conditions in Appendix A.2.1).

$$\mathbb{E}[\tilde{z}^{j'}_{i,t-s}\Delta\iota^{j}_{i,t}] = 0 \quad \text{for } t = 2, ..., T \text{ and } s \geq 1 \tag{1.11}$$

Using the Generalized Method of Moments, this modified Arellano-Bond estimator provides unbiased estimates of the productivity parameters $\{\alpha_1, \alpha_2, \alpha_3\}$.

Given the estimated productivity parameters $\{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3\}$, the TFP shock and production shock can be identified. First, compute the unobserved residual of equation (1.10) $r^j_{i,t} = \ln A_i + \eta_{i,t} + \tilde{\upsilon}^j_{i,t+1} - \alpha_1\tilde{\upsilon}^j_{i,t}$.

$$\hat{r}^j_{i,t} = \tilde{z}^j_{i,t+1} - \hat{\alpha}_1\tilde{z}^j_{i,t} - \hat{\alpha}_2 \ln L_{i,t} - \hat{\alpha}_3 \ln M_{i,t}$$

The variance of permanent TFP shock $\mathbb{V}(\ln A_i)$ is identified by the covariance of residuals associated with different measures in different periods, as production shocks are independent across periods and measurement errors are independent across measures.

$$Cov(r^j_{i,t}, r^{j'}_{i,t'}) = \mathbb{V}(\ln A_i) \quad \text{for } t \neq t' \text{ and } j \neq j'$$

The variance of temporary production shock $\mathbb{V}(\eta_{i,t})$ is identified by the covariance of residuals associated with different measures in the same period net of the variance of TFP shock $\mathbb{V}(\ln A_i)$.

$$Cov(r^j_{i,t}, r^{j'}_{i,t}) = \mathbb{V}(\ln A_i) + \mathbb{V}(\eta_{i,t}) \quad \text{for } j \neq j'$$

1.4.3 Preference Parameters

This section discusses the identification of child i's preference vector $\{\lambda_{i,t}, \tau\}$, parents i's preference vector $\{\delta_{i,t}, \xi_{i,t}\}$, and terminal valuations $\{\psi^c, \psi^p\}$.

Child's Preference

Child i's preference parameters in flow utility $\{\lambda_{i,t}, \tau\}$ are identified from the observed child effort $L_{i,t}$ by inverting the child's policy function (1.6).

$$\lambda_{i,t}(1 - \tau S_{i,t}) = \beta^c \Gamma^c_{t+1} \alpha_2 \frac{\bar{L} - L_{i,t}}{L_{i,t}} \tag{1.12}$$

where $\Gamma^c_t = 1 + \beta^c \Gamma^c_{t+1} \alpha_1$ and $\Gamma^c_{T+1} = \psi^c$. Let $R_{i,t} = \beta^c \Gamma^c_{t+1} \alpha_2 \frac{\bar{L} - L_{i,t}}{L_{i,t}}$. Conditional on $\{\beta^c, \psi^c, \alpha_1, \alpha_2\}$, given observed child effort $L_{i,t}$, $\lambda_{i,t}(1 - \tau S_{i,t})$ is identified by $R_{i,t}$. Theoretically, $\lambda_{i,t}$ is identified from observed child effort when parents do not discuss performance and τ is identified by within-child variation in parental performance discussion.[47]

Parents' Preference

Similarly, parental preference for consumption $\delta_{i,t}$ is identified from the observed parental educational investment $M_{i,t}$ by inverting parents' policy function (1.7).

$$\delta_{i,t} = \beta^p \Gamma^p_{t+1} \alpha_3 \frac{Y_{i,t} - M_{i,t}}{M_{i,t}} \tag{1.13}$$

where $\Gamma^p_t = 1 + \beta^p \Gamma^p_{t+1} \alpha_1$ and $\Gamma^p_{T+1} = \psi^p$. Conditional on $\{\beta^p, \psi^p, \alpha_1, \alpha_3\}$, given observed parental educational investment $M_{i,t}$, $\delta_{i,t}$ is identified by equation (1.13).

The log-normal distribution of parental performance discussion cost is characterized by a set of parameters $\Omega_\xi = \{\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \sigma_\xi\}$. The observed patterns of performance discussion choices are informative of Ω_ξ, as the distribution of performance discussion cost (along with the benefit of performance discussion) determines the probability of performance discussion.

$$q_{i,t} = \Phi\left(\frac{\ln B_{i,t} - \ln \xi_i}{\sigma_\xi}\right)$$

[47] Parental performance discussion decision $S_{i,t}$ is observed: when parents do not discuss performance ($S_{i,t} = 0$), $\lambda_{i,t}$ is identified by $R_{i,t}$; when parents discuss performance ($S_{i,t} = 1$), $\lambda_{i,t}(1 - \tau)$ is identified by $R_{i,t}$.

where $B_{i,t}$ is the benefit from performance discussion that can be calculated with other model parameters

$$B_{i,t} = \beta^p \Gamma^p_{t+1} \alpha_2 \ln \frac{\lambda_i + \beta^c \Gamma^c_{t+1} \alpha_2}{\lambda_i(1-\tau) + \beta^c \Gamma^c_{t+1} \alpha_2}$$

and performance discussion cost ξ_i is a parameterized function of consumption preference and observed household characteristics.

$$\xi_i = \pi_0 + \pi_1 \delta_i + \pi_2 \mathbf{1}\{x_{1i} \geq \text{college}\} + \pi_3 \mathbf{1}\{x_{2i} = \text{first child}\} + \pi_4 \mathbf{1}\{x_{3i} = \text{girl}\}$$

Conditional on the other parameters (that define $B_{i,t}$), the conditional probabilities of performance discussion by household traits $\mathbb{E}_t(q_{i,t}|x_{1i}, x_{2i}, x_{3i})$, the correlation of performance discussion and parental investment $\rho_t(S_{i,t}, M_{i,t})$, the distribution of total numbers of performance discussion in T periods $\mathbb{P}(J_i^S)$[48], and the probabilities of switching across performance discussion states $\{\mathbb{P}(S_{i,t+1} = 1|S_{i,t} = 0), \mathbb{P}(S_{i,t+1} = 0|S_{i,t} = 1)\}$[49] are informative about Ω_ξ that characterizes the distribution of performance discussion cost.

Terminal Valuation

Child and parental terminal valuations for final skill $\{\psi^c, \psi^p\}$ govern the time patterns of their investments. Thus, the moments of child effort $L_{i,t}$ and of parental monetary investment $M_{i,t}$ in each period are informative about ψ^c and ψ^p, respectively.

1.4.4 Estimation Algorithm

The estimation algorithm has two main steps. In the first step, I estimate the measurement parameters and production parameters directly with observed skill measures and investment choices without jointly solving the household behavior model. In the second

[48] The distribution of total numbers of performance discussion in T periods $\mathbb{P}(J_i^S)$ represents the within-individual variation in performance discussion that depends on σ_ξ. The intuition is that if $\sigma_\xi \to +\infty$, J_i^S should follow a binomial distribution with $n = T$ and $p = 0.5$, i.e., symmetric distribution with mode and center at $\frac{T}{2}$.

[49] The switching probabilities also reflect the within-individual variation in performance discussion that depends on σ_ξ.

step, I estimate the preference parameters by Simulated Method of Moments (see details in Appendix A.2.1). I target six sets of moments:

1. mean and standard deviation of percent of income spending on educational investment in each period: $\mathbb{E}_t(\frac{M_{i,t}}{Y_{i,t}})$, $\sigma_t(\frac{M_{i,t}}{Y_{i,t}})$
2. mean and standard deviation of child effort in each period: $\mathbb{E}_t(L_{i,t})$, $\sigma_t(L_{i,t})$
3. fraction of parents who discuss performance with their children by demographics in each period: $\mathbb{E}_t(q_{i,t}|x_{1i}, x_{2i}, x_{3i})$
4. correlation between performance discussion and percent of income spending on educational investment in each period: $\rho_t(S_{i,t}, \frac{M_{i,t}}{Y_{i,t}})$
5. probability of switching performance discussion status: $\mathbb{P}(S_{i,t+1} = 1|S_{i,t} = 0)$, $\mathbb{P}(S_{i,t+1} = 0|S_{i,t} = 1)$
6. distribution of total numbers of performance discussion in T periods: $\mathbb{P}(J_i^S)$

I conduct inference using non-parametric bootstrap. I re-sample the data at household level with replacement. For every bootstrapped sample, I re-estimate all parameters at each step in sequence, which ensures that the standard errors accurately reflect the sampling variation at each step of the estimator.

1.5 Model Estimates

1.5.1 Measurement Model of Latent Skill

Table 1.4 presents the estimated measurement parameters in equation 1.9 for three measures of latent skill in all periods. In the first period, the location and scale of each measure are determined by the initial normalization: logarithm of initial skill is normalized to be mean of 0 and standard deviation of 1. Math test score is an age-invariant measure for cognitive skill and its location and scale are constant over time. The locations and scales of other measures in later periods are identified as described in Section 1.4.1. The

Table 1.4: Measurement Parameters for Cognitive Skill

Measures	Location μ	Scale ω	%Signal	%Noise
Grade 5				
Math	0.110	0.644	0.518	0.482
English	0.136	0.705	0.610	0.390
Korean	0.081	0.584	0.445	0.555
Grade 6				
Math	0.110	0.644	0.580	0.420
English	0.270	0.679	0.632	0.368
Korean	0.026	0.575	0.567	0.433
Grade 7				
Math	0.110	0.644	0.639	0.361
English	0.315	0.639	0.755	0.245
Korean	0.102	0.524	0.543	0.457
Grade 8				
Math	0.110	0.644	0.612	0.388
English	0.495	0.699	0.729	0.271
Korean	0.057	0.594	0.631	0.369
Grade 9				
Math	0.110	0.644	0.615	0.385
English	0.262	0.742	0.747	0.253
Korean	0.219	0.644	0.647	0.353

Notes: The variance of $\ln\theta_{i,t}$ is estimated from the covariance of two different measures:

$$\mathbb{V}(\ln\theta_{i,t}) = \frac{Cov(z^j_{i,t}, z^{j'}_{i,t})}{\omega^j_t \omega^{j'}_t}$$

The percentage of signal is the faction of the variance of the measure that is explained by the latent skill:

$$\%Signal = \frac{(\omega^j_t)^2 \mathbb{V}(\ln\theta_{i,t})}{\mathbb{V}(z^j_{i,t})}$$

The percentage of noise is the faction of the variance of the measure that is due to measurement error:

$$\%Noise = \frac{\mathbb{V}(\varepsilon^j_{i,t})}{\mathbb{V}(z^j_{i,t})}$$

location of 0.11 for math score indicates that children with log skill of 0 score 0.11 on the math test on average. The scale of 0.64 is interpreted as an average increase of 0.64 in math score due to 1 standard deviation increase in latent log skill given that the standard deviation of log skill is normalized to be 1.

The last two columns in table 1.4 show the signal-and-noise variance decomposition for the measures. The percentage of signal in one measure is the fraction of observed variation in the measure that is associated with the latent skill. Signal ratios in these measures range from 45% to 75%. A higher percentage of signal suggests a more informative measure for the latent skill; as children grow older, all measures tend to have a higher information content.[50] The rest unexplained variance comes from measurement errors, referred to as percentage of noise, ranging from 25% to 55%. The evidence of non-trivial noises in all measures is the main motivation for correcting measurement errors when estimating the skill technology.

Based on the estimated measurement parameters, I depict the estimated development path of mean logarithm of cognitive skill in figure 1.3. As expected, I find that children's average latent (log) skill grows substantially by nearly 2 standard deviations from fifth to ninth grade. The growth rate in middle school (seventh to ninth grade) appears to be greater than that in elementary school (fifth to sixth grade).

1.5.2 Skill Technology

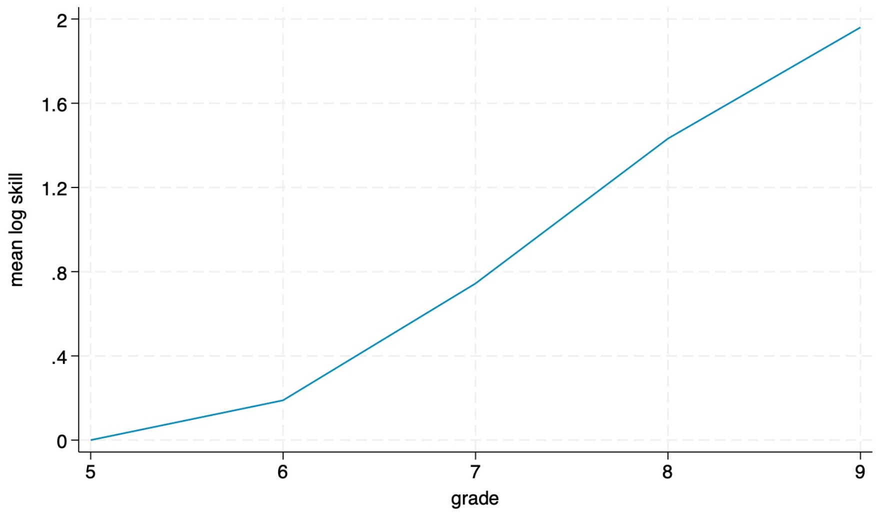
Panel A in table 1.5 reports the estimated production parameters that characterize the skill technology.[51] I find a persistence in skill level.[52] In addition, household educational investments - child effort and parental monetary investment - produce skills. Given the log-log functional form, production parameters α's can be interpreted as elasticity. Skill is

[50] The finding of higher signal ratios in measures at older ages is consistent with the existing literature (Cunha et al., 2010).

[51] When estimating the skill technology, the measurement errors in empirical measures for latent skill are corrected via the method described in section 1.4.2.

[52] The persistence parameter α_1 is lower than the estimates commonly found in classical value-added models that ignore individual heterogeneity in learning efficiency. As discussed thoroughly in Andrabi et al. (2011), this unobserved heterogeneity enters in every period and creates upward bias in estimated persistence in skill if ignored.

Figure 1.3: The Estimated Mean of Log Cognitive Skill



Notes: This figure shows the mean log skill (corrected for measurement error; source: KELS). The x-axis refers to the grade of children. Log skill at fifth grade is normalized to be mean 0.

more elastic in child effort than in parental investment. 1 percent increase in child effort increases skill by 0.23 percent, while 1 percent increase in parental investment raises it by 0.11 percent. I also investigate the empirical importance of variability in educational investments to skill formation based on the estimated productivity parameters and the empirical distribution of investments. I find that moving from the 10th to 90th percentile of the distribution of child effort or parental investment in each period increases produced skill by 15.1-19.6 percent or 15.2-16.9 percent.

Apart from observed inputs, skill development is also affected by unobserved shocks. The unobserved heterogeneity in children's learning efficiency (TFP) is substantial, accounting for the majority of unobserved variability in skill production. The estimated standard deviation of permanent (log) TFP shock is 1.1 units in standard deviation of latent skill, which is nearly double the standard deviation of temporary production shock. Failure to addressing the sizeable variability in TFP would lead to significant upward biases in estimation of productivity parameters, as part of the persistence in skill results from the individual-specific TFP that enters in every period and investments are found to be positively correlated with TFP.

Table 1.5: Estimates of Parameters

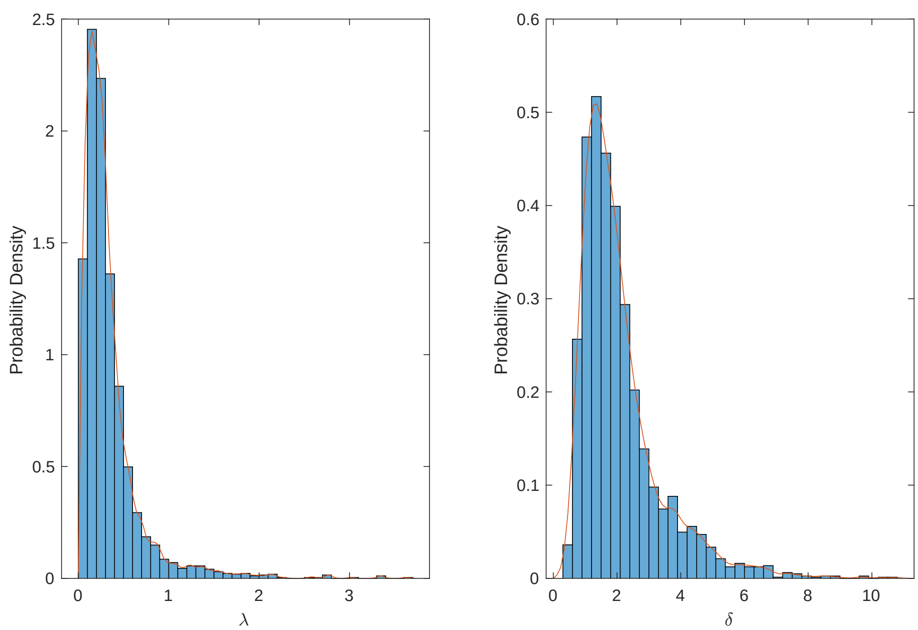
	Parameter	Estimate	95% CI
Panel A: Production Parameters			
self-productivity of skill	α_1	0.514	[0.313, 0.715]
productivity of child effort	α_2	0.233	[0.109, 0.351]
productivity of parental investment	α_3	0.110	[0.036, 0.179]
permanent TFP shock	$Var(\ln A_i)$	1.193	[0.841, 1.669]
temporary production shock	$Var(\eta_{i,t})$	0.286	[0.248, 0.329]
Panel B: Preference Parameters			
parents' preference for consumption	mean of δ_i	1.545	[0.476, 2.506]
	sd of δ_i	0.943	[0.285, 1.529]
child's preference for leisure	mean of λ_i	0.207	[0.105, 0.341]
	sd of λ_i	0.233	[0.115, 0.389]
performance discussion cost parameters	Ω_ξ		
mean cost:			
intercept	π_0	-2.973	[-3.682, -2.460]
slope of consumption preference	π_1	0.073	[0.028, 0.205]
slope of college-educated mother	π_2	-0.342	[-0.422, -0.280]
slope of first-born	π_3	-0.150	[-0.237, -0.069]
slope of girl	π_4	0.118	[0.041, 0.196]
standard deviation of cost	σ_ξ	0.545	[0.495, 0.616]
disutility from discussion × leisure	τ	0.549	[0.503, 0.590]
parent's discount factor	β^p	0.950	-
child's discount factor	β^c	0.950	-
parents' terminal value	ψ^p	1.037	[0.858, 1.258]
child's terminal value	ψ^c	1.045	[0.847, 1.350]

Notes: 95% confidence intervals are computed using bootstrap sampling household with replacement.

1.5.3 Behavioral Model

The behavioral model of parent-child interaction is estimated in the second stage after obtaining the productivity estimates. Panel B of table 1.5 presents the estimates of preference parameters. The model allows for rich across-population heterogeneity in preferences. I measure the dispersion of preferences by the ratio of standard deviation to mean (i.e., the coefficient of variation), which is 0.610 for parents' preference of consumption and 1.126 for children's preference of leisure, respectively. The larger dispersion in child leisure preference is due to a longer right tail of the distribution shown in figure 1.4. The finding of some extremely higher preference for leisure reflects the empirical fact that some children exert minimal effort in class. Discussion cost also varies across population. I find a positive correlation of discussion cost and consumption preference, indicating that parents who place a high value on child skill tend to have relatively low consumption preference and low discussion cost. In addition, discussion cost differs by maternal education, birth order, and child gender. The lower costs of highly-educated mothers can be interpreted as better communication skills with children. Parents bear lower costs to discipline the first-born child if they benefit additionally from deterring undesirable behaviors of the later-born by being strict to the oldest.[53] The significant gender difference in discussion cost suggests that parenting strategies are tailored to the gender of child. Lastly, the standard deviation of discussion cost captures the within-household variation in discussion probability over time.

Figure 1.4: Distributions of Heterogeneous Preferences



Notes: λ_i is children's preference for leisure; δ_i is parents' preference for consumption (Source: KELS)

Other parameters are assumed to be common across households. The parameter that

[53]As discussed in section 1.2.4, discussion cost is modeled in a reduced-form way in that it is a combination of various sources of parental utility associated with discussion.

governs the disutility from the interaction of performance discussion and leisure is estimated to be 0.549, which implies that their utility from leisure is reduced by more than a half when having performance discussion. I fix parents' and children's annualized discount factors at 0.95. The estimated terminal valuations of child final skill for parents and children are 1.037 and 1.045, respectively.

1.5.4 Joint Distribution of Unobserved Heterogeneity

Table 1.6 reports the estimated correlation matrix of unobserved heterogeneity. As discussed in the model section, households are heterogeneous in four aspects including child learning efficiency, child preference for leisure, parental preference for consumption, and performance discussion cost. The consideration of flexible joint distribution of unobserved heterogeneity "corrects" for the endogeneity of inputs in the estimation of skill technology. I find a significantly negative correlation of child learning efficiency and parents' preference for consumption, which explains the positive relationship between learning efficiency and parental education investment found in the skill technology. Similarly, the significantly negative correlation of child learning efficiency and their preference for leisure indicates a positive relationship between learning efficiency and effort. If not taking into account these correlations, the estimated productivity of child effort and parental investment would be positively biased, because children who make more effort due to lower leisure preference and/or receive more parental investment due to lower parental consumption preference tend to have higher learning efficiency that directly leads to higher skill. In addition, performance discussion cost is negatively correlated with child learning efficiency and positively correlated with child leisure preference, suggesting that parents' choice of performance discussion depends on child unobserved traits.

1.5.5 Sample Fit

Table 1.7 displays the sample fit of the simulated model at estimated parameters to the key general patterns of household behaviors. The model fits well the distribution of

Table 1.6: Correlations of Unobserved Heterogeneity

Correlation		learning ability A_i	child leisure λ_i	parental consumption δ_i
child leisure	λ_i	-0.111		
		[-0.161, -0.065]		
parental consumption	δ_i	-0.050	0.008	
		[-0.096, -0.009]	[-0.033, 0.048]	
parental discussion cost	ξ_i	-0.131	0.408	0.064
		[-0.189, -0.084]	[0.170, 0.736]	[0.020, 0.100]

Notes: 95% confidence intervals in brackets are computed using bootstrap sampling household with replacement.

educational investments. It accurately predicts the mean of child effort and the mean fraction of household income spending on educational investment in each period, though slightly underestimates their standard deviations. Moreover, the estimated model is able to replicate the correlation of two parental choices - performance discussion and propensity of investment in child education - and the within-individual variability in performance discussion that reflects in the distribution of total numbers of discussion in all periods and the probability of switching discussion states across periods.

Besides the overall features of household behaviors, table 1.8 presents the sample fit of performance discussion decisions by subgroups. The model reproduces the differential rates of performance discussion by maternal education, birth order, and child gender. It captures the pattern of higher discussion rates when the mother is college-educated, the child is first born, and/or the child is a boy.

Table 1.7: Sample Fit of Child Effort, Parental Investment, and Performance Discussion

(a) Child Effort

	Mean		SD	
Period	Data	Model	Data	Model
$t = 1$	0.6984	0.7093	0.1651	0.1685
$t = 2$	0.7044	0.7121	0.1853	0.1722
$t = 3$	0.6919	0.7105	0.1935	0.1735
$t = 4$	0.6772	0.6945	0.2039	0.1786

(b) Fraction of Income Spending on Education

	Mean		SD	
Period	Data	Model	Data	Model
$t = 1$	0.0937	0.9036	0.0616	0.0486
$t = 2$	0.1035	0.9033	0.0603	0.0485
$t = 3$	0.1066	0.0921	0.0654	0.0480
$t = 4$	0.1084	0.0852	0.0626	0.0449

(c) Performance Discussion and Frac. of Income Spending on Educ.

	Correlation	
Period	Data	Model
$t = 1$	0.5338	0.5527
$t = 2$	0.5629	0.5522
$t = 3$	0.5640	0.5496
$t = 4$	0.5288	0.5349

(d) Within-Individual Variation in Performance Discussion

Distribution of No. of Discussion			Switching Probability	
$\mathbb{P}(J_i^S)$	Data	Model	Data	Model
$J_i^S = 0$	0.1849	0.2054	$\mathbb{P}(S_{i,t+1} = 0 \mid S_{i,t} = 1)$	
$J_i^S = 1$	0.1722	0.1593	0.1400	0.1449
$J_i^S = 2$	0.1678	0.1766	$\mathbb{P}(S_{i,t+1} = 1 \mid S_{i,t} = 0)$	
$J_i^S = 3$	0.2072	0.2101	0.1390	0.1578
$J_i^S = 4$	0.2679	0.2486		

Notes: Data columns show the actual data moments. Model columns show the model prediction at estimated parameters.

Table 1.8: Sample Fit of Conditional Probability of Performance Discussion

(a) Low-Education Mothers

	First-born				Later-born			
	Boy		Girl		Boy		Girl	
Period	Data	Model	Data	Model	Data	Model	Data	Model
$t = 1$	0.6116	0.5735	0.4979	0.4632	0.5325	0.5090	0.4390	0.3962
$t = 2$	0.5446	0.5744	0.5187	0.4641	0.5232	0.5100	0.4884	0.3972
$t = 3$	0.5848	0.5799	0.4647	0.4701	0.4954	0.5158	0.4419	0.4033
$t = 4$	0.5536	0.6115	0.4855	0.5050	0.4644	0.5497	0.3953	0.4400

(b) High-Education Mothers

	First-born				Later-born			
	Boy		Girl		Boy		Girl	
Period	Data	Model	Data	Model	Data	Model	Data	Model
$t = 1$	0.5879	0.6304	0.5701	0.5407	0.5440	0.5682	0.5190	0.4912
$t = 2$	0.6571	0.6312	0.5986	0.5416	0.6000	0.5690	0.5831	0.4921
$t = 3$	0.6888	0.6362	0.6295	0.5468	0.5920	0.5741	0.6122	0.4977
$t = 4$	0.6369	0.6648	0.6081	0.5774	0.5760	0.6042	0.5452	0.5306

Notes: Data columns show the actual data moments. Model columns show the model prediction at estimated parameters.

1.6 Quantify Indirect Parental Impact from Performance Discussion

In this paper, the key innovation of the model is that parents can influence children's skills in a novel indirect channel where they engage in performance discussion to affect children's effort choices, besides a standard direct channel where they make monetary investment in children's education. I examine the importance of indirect parental influence through the discussion channel to children's skill formation from two perspectives. In terms of average skill development, I estimate the impact of performance discussion on child effort and benchmark the relative importance of performance discussion to improving skill with that of parental monetary investment. Moreover, I discuss how differential likelihoods of performance discussion contribute to the skill gap between children from high- and low-SES families and to what extent closing the discussion gap slows down the expansion of the skill gap during adolescence.

1.6.1 Average Contribution of Performance Discussion to Child Skill

In this section, I first quantify the importance of performance discussion to child effort by comparing average child effort predicted by the estimated baseline model and by a counterfactual model that shuts down the performance discussion channel. Based on the simulated investment choices, I simulate and compare the average skill development paths in the two models to estimate the contribution of performance discussion to skill formation. Moreover, I investigate the potential impact of promoting performance discussion on skill development, forming the basis of forecasting the effectiveness of policy interventions encouraging parental involvement in child schooling (e.g., persuasive communication) to improve skill.

Table 1.9: Contribution of Performance Discussion to Child Effort

Model Period	(1) Baseline Model	(2) No Discussion	(3) Decrease (p.p.)	(4) Decrease (%)
$t = 1$	70.35	62.93	-7.42	-10.55
	(0.34)	(0.47)	(0.48)	(0.68)
$t = 2$	65.28	57.37	-7.91	-12.12
	(0.31)	(0.47)	(0.50)	(0.70)
$t = 3$	64.13	56.23	-7.90	-12.32
	(0.34)	(0.46)	(0.50)	(0.70)
$t = 4$	62.78	55.28	-7.50	-11.95
	(0.37)	(0.57)	(0.49)	(0.68)

Notes: Standard errors in parentheses computed using bootstrap sampling households with replacement.
Column (1): average child effort measured by percentage of time paying attention in class in baseline model
Column (2): average child effort measured by percentage of time paying attention in class in the counterfactual model that shuts down the performance discussion channel by setting τ to be 0
Column (3): the percentage-point change in average child effort from baseline to the no-discussion counterfactual
Column (4): the percent change in average child effort from baseline to the no-discussion counterfactual

Precluding Performance Discussion Decreases Child Effort and Inhibits Skill Development

I quantify the average contribution of performance discussion to child effort by simulating the counterfactual child effort when the discussion channel shuts down[54] and comparing it to the baseline effort. Table 1.9 shows that if parents are unable to influence their child's effort choice by discussion, average effort decreases by 7.42-7.91 percentage points (or 10.55-12.32 percent) from the baseline. In other words, 7.42-7.91 p.p. (or 10.55-12.32 percent) of average child effort in the baseline is attributed to the impact of performance discussion.

Given the simulated child effort choices when the discussion channel shuts down, I

[54] To shut down the discussion channel, I set the parameter τ that governs the disutility from the interaction of performance discussion and leisure to be 0 so that no parents would discuss performance given no impact of discussion on effort. I simulate the investment choices at the estimated parameters (except τ) in the counterfactual model and compute the average child effort.

Table 1.10: Contribution of Performance Discussion to Child Skill Development

Model Period t	(1) Baseline Model $\ln \theta_{t+1}$	(2) Counter-factual $\ln \theta_{t+1}$	(3) Change in Skill $\Delta \ln \theta_{t+1}$	(4) Equiv. investment ΔI_t^p	(5) Equiv. income ΔY_t
Panel (A): Counterfactual 1 - No Discussion					
1	0.200	0.155	-0.044	\$1,006	\$10,537
2	0.751	0.676	-0.075	\$1,303	\$14,092
3	1.406	1.322	-0.084	\$1,407	\$16,778
4	1.924	1.834	-0.090	\$1,424	\$21,798
Panel (B): Counterfactual 2 - All Discussion					
1	0.200	0.206	0.006	\$736	\$8,182
2	0.751	0.760	0.009	\$934	\$10,639
3	1.406	1.436	0.030	\$1,034	\$12,999
4	1.924	1.964	0.040	\$1,217	\$19,971

Notes: Counterfactual 1 - No Discussion: I shut down the discussion channel by setting the parameter that governs the change in the child's value for leisure due to discussion to be 0 so that no parents discuss performance given no impact from discussion on child effort choices.
Counterfactual 2 - All Discussion: To facilitate discussion from all parents, I generate the counterfactual discussion cost for each household in every period such that it is smaller than the discussion benefit.
Column (1): average (log) skill at the end of period t in baseline model
Column (2): average (log) skill at the end of period t in counterfactual model. I first simulate children's effort choices given counterfactual discussion decisions and then calculate the corresponding skill development paths with the estimated skill technology.
Column (3): average change in standard deviation of (log) skill at the end of period t from baseline to counterfactual
Column (4): conversion of the change in column (3) into an equivalent change in parental investment in period t
Column (5): conversion of the change in column (3) into an equivalent change in household income in period t (holding constant household investment propensity)
The conversion in columns (4) and (5) is done in the following way. First, holding constant discussion behaviors, how much reduction/rise in parental monetary investment is needed to rationalize the decrease/increase in skill. Then, given household investment propensity, how much reduction/rise in household income is needed to rationalize the reduction/rise in educational investment.

further simulate the corresponding skill development path according to the estimated skill technology. Panel (A) in table 1.10 reports the comparison between the baseline skill development path in column (1) and the counterfactual outcome when the discussion channel shuts down in column (2). I find that if parents are unable to influence child effort through performance discussion, average skill at ninth grade decreases by around 0.090 standard deviation from the baseline. To benchmark the relative importance of performance discussion to skill development with parental monetary investment, I convert the discussion effect on skill into monetary terms:[55] shutting down the channel of performance discussion is equivalent to an average of approximately \$5,140 reduction in educational investment in four years or a corresponding \$63,206 reduction in household income (given the estimated investment propensity).

Table 1.11: Overall Parental Influence on Skill Development (Direct and Indirect Channels)

Baseline four-year skill development	1.924 SD
Counterfactual 1: shut down the performance discussion channel	
Impact on skill development	↓ 0.090 SD
Counterfactual 2: shut down the monetary investment channel	
Impact on skill development	↓ 0.222 SD
Overall parental influence on skill development	0.312 SD
from indirect channel	29.0%
from direct channel	71.0%

Notes: Counterfactual 1 - No Discussion: I shut down the discussion channel by setting the parameter that governs the change in the child's value for leisure due to discussion to be 0 so that no parents discuss performance given no impact from discussion on child effort choices.
Counterfactual 2 - No Monetary Investment: I shut down the monetary investment channel by setting the productivity parameter of monetary investment to be 0 so that no parents make monetary investment given no impact on child skill.

Another way to quantify the importance of performance discussion is to decompose the overall parental influence on skill development into the two channels. Table 1.11 shows that 29 percent of overall parental influence on four-year skill development is attributed to the performance discussion channel. In other words, neglecting the indirect impact from performance discussion will underestimate overall parental influence by 29 percent.

[55] The conversion is done in the following way. First, if holding constant discussion behaviors, how much reduction in parental monetary investment is needed to rationalize the decrease in skill. Then, given household investment propensity, how much reduction in household income is needed to rationalize the reduction in educational investment.

Promoting Performance Discussion Increases Child Effort and Improves Skill Development

In the baseline model, not all parents choose to discuss performance, which suggests that promoting performance discussion in the population could improve skill formation of children whose parents never or rarely discuss performance. Table 1.9 shows that if all parents choose to influence their child's effort choice by discussion,[56] average effort increases by around 4.68-5.40 percentage points (or 6.65-8.60 percent) from the baseline. Panel (B) in table 1.10 compares the average skill development path in the counterfactual model that facilitates performance discussion for all parents to the baseline model. The predicted increase of 0.040 standard deviation in final skill is the upper bound of the potential impact of promoting performance discussion, as the best possible outcome is that all parents discuss performance in all periods due to the promotion. I benchmark the impact of promoting discussion on skill formation by converting it into monetary terms: if all parents are promoted to discuss performance with their child, it is equivalent to increasing four-year parental educational investment by \$3,922 or to giving (unconditional) income transfer of \$51,791.

1.6.2 Heterogeneous Discussion Effects, Discussion Gap, and Skill Gap

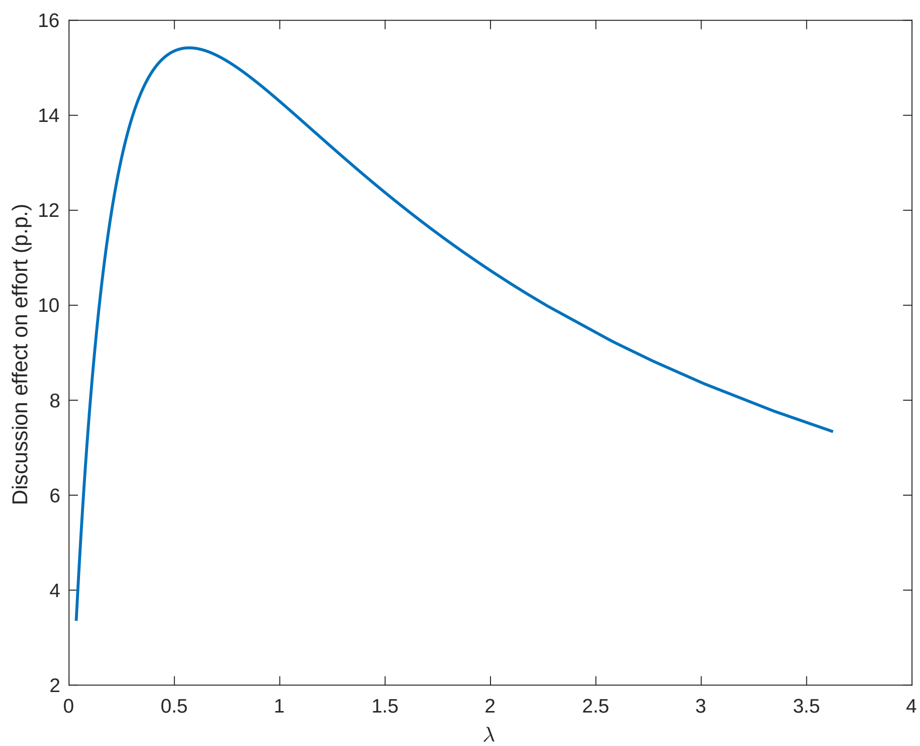
In this section, I first describe the heterogeneity in the effect of performance discussion on child effort. Then I discuss how heterogeneous discussion effects lead to heterogeneous benefits from discussion, which explain differential discussion behaviors across population along with variability in discussion costs. Furthermore, I explore if addressing discussion gap is an effective way to closing skill gap between socioeconomically advantaged and disadvantaged groups.

[56] To facilitate discussion for all parents, I generate the counterfactual discussion cost for each household in every period such that it is smaller than the discussion benefit. I simulate children's effort choices given discussion from parents and calculate the corresponding skill development paths with the estimated skill technology.

Heterogeneous Effects of Performance Discussion on Effort

The model restricts the parameter that governs the disutility from the interaction of performance discussion and leisure to be common across population, but it is not equivalent to assuming homogeneous effect of performance discussion on child effort. Instead, the effect of discussion on effort is allowed to vary with children's preference for leisure,[57] because of concave marginal utility from leisure (or equivalently, convex marginal cost of effort) and diminishing returns of child effort in skill production, which will be discussed further in the next paragraph. Figure 1.5 plots this heterogeneity in discussion effect by

Figure 1.5: Heterogeneous Effects of Performance Discussion on Child Effort



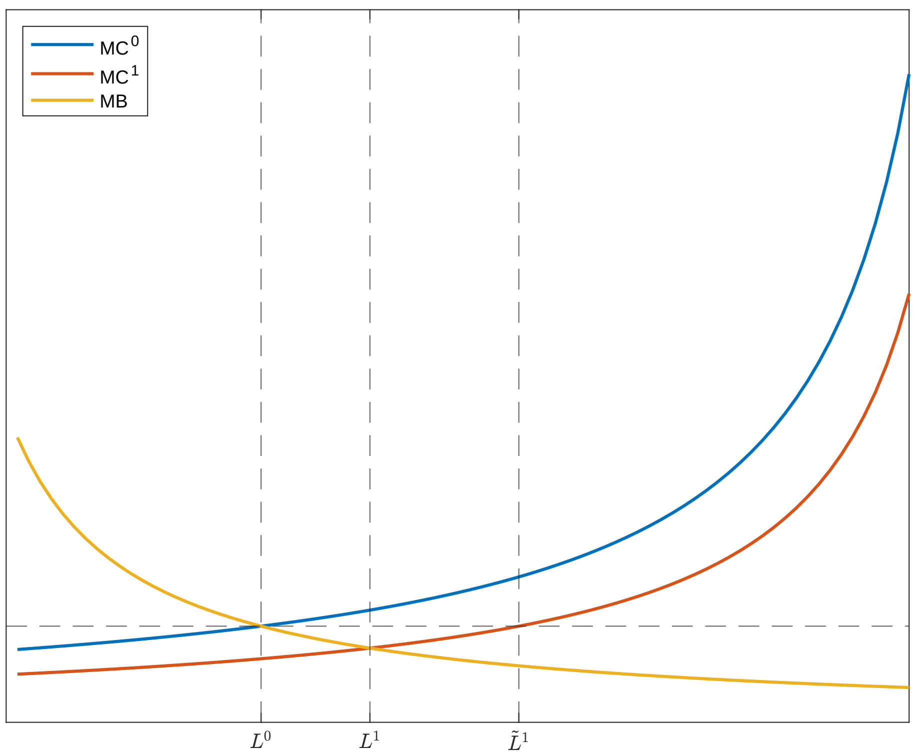
Notes: In each period t, for child i, there are two potential outcomes for $L^*_{i,t}(S_{i,t})$ depending on $S_{i,t}$. Let $L^1_{i,t} = L^*_{i,t}(S_{i,t} = 1) = \frac{\bar{L}\beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_{i,t}(1-\tau)+\beta^c\Gamma^c_{t+1}\alpha_2}$ and $L^0_{i,t} = L^*_{i,t}(S_{i,t} = 0) = \frac{\bar{L}\beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_{i,t}+\beta^c\Gamma^c_{t+1}\alpha_2}$. The figure plots $(L^1_{i,t} - L^0_{i,t})$ by $\lambda_{i,t}$.

child leisure preference (λ_i). The impact of performance discussion is nonlinear in leisure preference: children's responsiveness to discussion first increases and then decreases in

[57]Leisure preference is an inverse measure of diligence as a lower leisure preference means a more diligent child who makes higher effort in the absence of discussion.

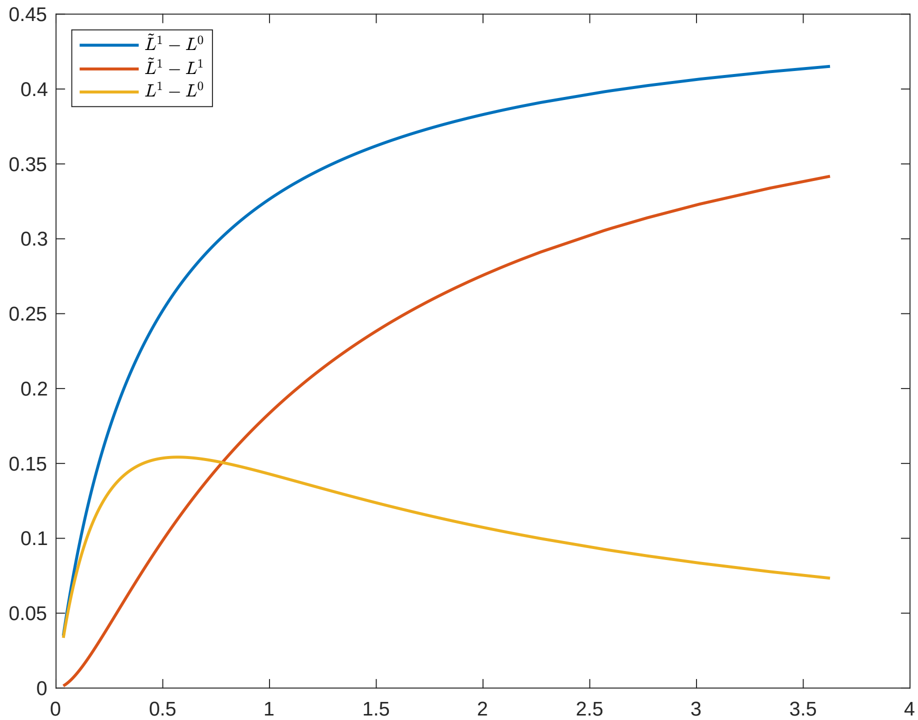
preference for leisure. In the left region of low leisure preference $\lambda_i \in (0, 0.3)$, the lower value the child places on leisure, the smaller the increase in effort becomes. Intuitively, the most diligent children (characterized by the lowest leisure preference) already exert high effort and it is very costly to improve further due to the convexity of cost. The most responsive children are those with medium preference for leisure $\lambda_i \in (0.3, 0.4)$, raising effort by around 19 percentage points. In the right region of high leisure preference $\lambda_i \in (0.4, 2.0)$, the effect of discussion decreases in preference for leisure.

Figure 1.6: Mechanism of Performance Discussion on Child Effort



Notes: This figure illustrates how children choose effort optimally by equating marginal cost and marginal benefit of effort.
MB: the curve of marginal benefit of effort $\beta^c \Gamma^c_{t+1} \frac{\alpha_2}{L_{i,t}}$
MC^0: the curve of marginal cost of effort in the absence of discussion $\frac{\lambda_{i,t}}{\bar{L} - L_{i,t}}$
MC^1: the curve of marginal cost of effort in the presence of discussion $\frac{\lambda_{i,t}(1-\tau)}{\bar{L} - L_{i,t}}$
The optimal choice of effort without discussion $L^0_{i,t} = L^*_{i,t}(S_{i,t} = 0) = \frac{\bar{L}\beta^c \Gamma^c_{t+1} \alpha_2}{\lambda_{i,t} + \beta^c \Gamma^c_{t+1} \alpha_2}$ is pinned down by the intersection of MB and MC^0.
The optimal choice of effort with discussion $L^1_{i,t} = L^*_{i,t}(S_{i,t} = 1) = \frac{\bar{L}\beta^c \Gamma^c_{t+1} \alpha_2}{\lambda_{i,t}(1-\tau) + \beta^c \Gamma^c_{t+1} \alpha_2}$ is pinned down by the intersection of MB and MC^1.
If the marginal benefit of effort is constant (a flat MB curve), the counterfactual choice of effort with discussion is $\tilde{L}^1_{i,t} = \bar{L} - \frac{\lambda_{i,t}(1-\tau)}{\beta^c \Gamma^c_{t+1} \alpha_2} L^0_{i,t}$.

Figure 1.7: Decomposition of Performance Discussion Effect on Child Effort



Notes: $L^0_{i,t} = L^*_{i,t}(S_{i,t}=0) = \frac{\bar{L}\beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_{i,t}+\beta^c\Gamma^c_{t+1}\alpha_2}$: the optimal choice of effort without discussion
$L^1_{i,t} = L^*_{i,t}(S_{i,t}=1) = \frac{\bar{L}\beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_{i,t}(1-\tau)+\beta^c\Gamma^c_{t+1}\alpha_2}$: the optimal choice of effort with discussion
$\tilde{L}^1_{i,t} = \bar{L} - \frac{\lambda_{i,t}(1-\tau)}{\beta^c\Gamma^c_{t+1}\alpha_2}L^0_{i,t}$: the counterfactual choice of effort with discussion holding constant the marginal benefit
The figure shows the decomposition of discussion effect on effort $L^1_{i,t} - L^0_{i,t} = (\tilde{L}^1_{i,t} - L^0_{i,t}) - (\tilde{L}^1_{i,t} - L^1_{i,t})$.
$\tilde{L}^1_{i,t} - L^0_{i,t} = \frac{\bar{L}\lambda_{i,t}\tau}{\lambda_{i,t}+\beta^c\Gamma^c_{t+1}\alpha_2}$: the counterfactual increase in effort due to reduced marginal cost (holding constant the marginal benefit)
$\tilde{L}^1_{i,t} - L^1_{i,t} = \frac{\bar{L}\lambda_{i,t}(1-\tau)}{\lambda_{i,t}(1-\tau)+\beta^c\Gamma^c_{t+1}\alpha_2}\frac{\lambda_{i,t}\tau}{\lambda_{i,t}+\beta^c\Gamma^c_{t+1}\alpha_2}$: the counterfactual decrease in effort due to diminishing marginal benefit of effort

To help make sense of the heterogeneous effects of discussion on child effort, I first illustrate the mechanism of discussion effect with an analysis of marginal cost and benefit of child effort in figure 1.6 and then show how the decomposition pattern of discussion effect varies by child leisure preference in figure 1.7. Performance discussion reduces marginal cost of child effort,[58] equivalent to shifting the curve of marginal cost outward as shown in figure 1.6. If marginal benefit of effort is constant, the counterfactual increase in effort is completely attributed to the change in marginal cost associated with discussion.

[58]The idea that discussion reduces the marginal cost of effort by decreasing child utility from leisure can be interpreted as discussion increases the child's relative value of skill over leisure.

However, marginal benefit of effort decreases in effort because of diminishing returns of effort in skill technology, which counteracts part of the increase from reduced cost. To see how the magnitude of the two sources differs by child leisure preference, figure 1.7 plots the decomposition of discussion effect against leisure preference into the counterfactual increase in effort due to reduced marginal cost holding constant marginal benefit and the counterfactual decrease in effort due to diminishing marginal benefit. In the region of low leisure preferences, the slope of the counterfactual increase is steeper than that of the counterfactual decrease, leading to a rising trend in actual change in effort. In contrast, the reverse is true for higher leisure preferences. Taking together, the two figures explain the non-linearity of discussion effect on effort in leisure preference.

Heterogeneous Probabilities of Performance Discussion

Performance discussion decision hinges on the trade-off of discussion cost in current period and benefit from a higher skill in future periods. The estimated model confirms two sources of heterogeneity in discussion probabilities. From the perspective of discussion benefit, its variation results from heterogeneous discussion effects on effort discussed in section 1.6.2. Loosely speaking, the benefit reflects the percentage increase in effort due to performance discussion, that is, the ratio of the actual increase in effort (measured by discussion effect on effort) over the effort in absence of discussion. The model predicts that discussion benefit monotonically increases in child leisure preference.[59] The other source of heterogeneity in discussion probabilities comes from variation in discussion costs by households' observed and unobserved characteristics as shown in section 1.5.3. Taking together, the model explains the observed differential probabilities of discussion across subgroups. First, the two sources work in opposite directions for the discussion gap between high- and low-socioeconomic groups.[60] High-SES parents receive lower discussion

[59]The declining discussion effect and the increasing discussion benefit in the region of high leisure preferences are reconciled in that although actual increase in effort is smaller for high-leisure-preference children than medium children, the former's effort in absence of discussion is also much lower, so the percentage increase is higher.

[60]I measure socioeconomic status (SES) by maternal education and define households with a college-educated mother as high SES.

benefits because of lower leisure preference of their children, but they have lower discussion costs.[61] I find that the effect of lower costs dominates that of lower benefits, resulting in higher discussion probability of high-SES parents. Similarly, for first-born children, lower discussion costs dominate lower discussion benefits associated with lower leisure preference. This rationalizes the phenomenon of strategic parenting that first-born children are more likely to have performance discussion with their parents, which is consistent with the finding in Hotz and Pantano (2015) that parents tend to impose more stringent disciplinary restrictions on the earlier-born in order to deter bad outcomes for the later-born. Moreover, I find that significant gender difference in discussion probability is due to higher discussion costs associated with girls[62] rather than different discussion benefits, because leisure preferences do not vary systematically by gender. Last but not least, the positive correlation of consumption preference and discussion cost and the insignificant correlation of consumption preference and leisure preference indicate that low-consumption-preference parents who have a higher propensity to spend on child education are also more likely to discuss performance because of lower discussion costs; in other words, these parents are consistently more actively involved in child learning process through multiple channels.

Addressing Performance Discussion Gap Closes Skill Gap

In terms of skill formation, how to improve skill on average is a critical issue. What is equally important is how to reduce the skill gap between advantaged and vulnerable groups. Performance discussion is a source of skill disparity between children from high- and low-socioeconomic groups and it affects the skill distribution for two reasons. First, section 1.6.2 illustrates that performance discussion has heterogeneous effects on child effort. Second, parents face heterogeneous discussion costs and benefits, leading to differential probabilities of discussion. Section 1.6.2 shows that high-SES parents are more likely to discuss performance with their children, not because of differential discussion benefits, but due to lower discussion costs.

[61] It may suggest that high-skilled parents have better communication skills with children.

[62] Higher costs may reflect that adolescent girls are less willing to communicate with their parents.

Table 1.12: Equalize Discussion Cost to Close Child Skill Gap

High- and Low-SES Skill Gap: $\mathbb{E}(\ln\theta \mid x_1 = 1) - \mathbb{E}(\ln\theta \mid x_1 = 0)$	
Initial Skill Gap	0.595
Expansion of Skill Gap (Baseline)	0.230
Expansion of Skill Gap (Counterfactual of Equal Discussion Cost)	0.212
Slowdown in Expansion of Skill Gap	-0.018
equiv. to increasing parental investment for low-SES ($)	1,491
$t = 1$	269
$t = 2$	355
$t = 3$	396
$t = 4$	472
equiv. to unconditional income transfer to low-SES ($)	19,431
$t = 1$	2,971
$t = 2$	4,034
$t = 3$	4,906
$t = 4$	7,520

Notes: I measure socioeconomic status (SES) by maternal education and define households with a college-educated mother ($x_1 = 1$) as high SES.
Expansion of Skill Gap (Baseline): the increase in skill gap from initial to final period in the baseline model
Expansion of Skill Gap (Counterfactual of Equal Discussion Cost): the increase in skill gap from initial to final period in the counterfactual model that equalizes average discussion cost across socioeconomic groups
Slowdown in Expansion of Skill Gap: how equalizing discussion cost across socioeconomic groups slows down the expansion of skill gap

Since it is the difference in discussion costs that explains the discussion gap across socioeconomic groups, I examine how addressing performance discussion gap by eliminating systematic difference in discussion cost might affect the skill gap. Table 1.12 presents the skill gap between children from high- and low-SES groups predicted by the baseline model and by a counterfactual model that reduces the discussion cost of low-SES parents to the level of the high-SES. The initial skill gap is substantial: average skill of high-SES children is 0.595 standard deviation higher than that of the low-SES. The baseline skill gap widens by 0.230 standard deviation in four years. When equalizing average discussion cost in high- and low-socioeconomic groups, the expansion in skill gap is closed by 0.018 standard deviation or 7.9 percent. To benchmark the significance of performance discussion to skill gap, I convert the impact of closing performance discussion gap into monetary terms: to slow down the expansion of skill gap, equalizing average discussion cost is equivalent to increasing educational investment of low-SES parents by \$1,491 or to giving them (unconditional) income transfer of \$19,431 in four years.

1.7 Bounding Analysis

1.7.1 Model Extension: Correcting Bias from Reverse Causality

Reverse Causality Causes Negative Bias

The identification of the effect of performance discussion on child effort relies on the model timing of parent-child interaction that parents choose whether to discuss performance before preference shock to child leisure is realized and the child chooses effort.[63] This assumption of parents' not knowing leisure/effort shocks ϵ^{λ} may be at odds with the data. Parents and children answer survey questions in a retrospective way: they report the average action in the past year, so the timing of actions is unclear. It is possible that parents have some knowledge of shocks to child effort in a given period and make choices of performance discussion accordingly. This raises the concern of reverse causality that shocks to child effort affect parents' discussion decision,[64] which potentially leads to bias in the estimated effect of performance discussion on child effort.

This section presents a formal discussion of why ignoring reverse causality would lead to a negative bias in estimated $\hat{\tau}$, the parameter that governs the effect of performance discussion on effort. First, I describe how the parameter τ is estimated in the baseline model. Second, I discuss how the baseline estimation of τ is negatively biased when I relax the assumption that parents do not observe preference shocks to child leisure and allow leisure shocks to affect parents' choice of performance discussion.

In the baseline model, the independence of performance discussion and leisure shocks,

[63] In the model, preference shock to child leisure is equivalent to shock to child effort.

[64] For example, when parents observe a negative shock to child effort, they may be more likely to take actions to influence child choices.

$S_{i,t} \perp \epsilon^{\lambda}_{i,t}$, indicates that

$$\mathbb{E}(\ln \epsilon^{\lambda}_{i,t} | S_{i,t}) = \mathbb{E}(\ln \epsilon^{\lambda}_{i,t}) = 0$$

Suppose that parents i discuss performance in period t^i_1 but not in period t^i_2, i.e., $S_{i,t^i_1} = 1$ and $S_{i,t^i_2} = 0$. Rearrange child policy function of effort in periods t^i_1 and t^i_2 and plug in $S_{i,t^i_1} = 1$ and $S_{i,t^i_2} = 0$.

$$\lambda_{i,t^i_1}(1-\tau) = \beta^c \Gamma^c_{t+1} \alpha_2 \frac{\bar{L} - L_{i,t^i_1}}{L_{i,t^i_1}} \tag{1.14}$$

$$\lambda_{i,t^i_2} = \beta^c \Gamma^c_{t+1} \alpha_2 \frac{\bar{L} - L_{i,t^i_2}}{L_{i,t^i_2}} \tag{1.15}$$

Divide equation (1.14) by equation (1.15) and rearrange.

$$\frac{\epsilon^{\lambda}_{i,t^i_1}}{\epsilon^{\lambda}_{i,t^i_2}}(1-\tau) = \frac{\bar{L} - L_{i,t^i_1}}{L_{i,t^i_1}} \frac{L_{i,t^i_2}}{\bar{L} - L_{i,t^i_2}}$$

$$\ln(1-\tau) = \ln\left(\frac{\bar{L} - L_{i,t^i_1}}{L_{i,t^i_1}} \frac{L_{i,t^i_2}}{\bar{L} - L_{i,t^i_2}}\right) + \ln \epsilon^{\lambda}_{i,t^i_2} - \ln \epsilon^{\lambda}_{i,t^i_1}$$

Suppose that there are N^τ parents who switch discussion status.

$$\ln(1-\tau) = \frac{1}{N^\tau}\sum_{i=1}^{N^\tau} \ln\left(\frac{\bar{L} - L_{i,t^i_1}}{L_{i,t^i_1}} \frac{L_{i,t^i_2}}{\bar{L} - L_{i,t^i_2}}\right) + \underbrace{\frac{1}{N^\tau}\sum_{i=1}^{N^\tau} \ln \epsilon^{\lambda}_{i,t^i_2}}_{\xrightarrow{p} \mathbb{E}(\ln \epsilon^{\lambda}_{i,t^i_2} | S_{i,t^i_2}=0) = \mathbb{E}(\ln \epsilon^{\lambda}_{i,t^i_2}) = 0}$$

$$- \underbrace{\frac{1}{N^\tau}\sum_{i=1}^{N^\tau} \ln \epsilon^{\lambda}_{i,t^i_1}}_{\xrightarrow{p} \mathbb{E}(\ln \epsilon^{\lambda}_{i,t^i_1} | S_{i,t^i_1}=1) = \mathbb{E}(\ln \epsilon^{\lambda}_{i,t^i_1}) = 0}$$

Equation (1.16) gives a consistent estimate of τ.

$$\ln(1-\hat{\tau}) = \frac{1}{N^\tau}\sum_{i=1}^{N^\tau} \ln\left(\frac{\bar{L} - L_{i,t^i_1}}{L_{i,t^i_1}} \frac{L_{i,t^i_2}}{\bar{L} - L_{i,t^i_2}}\right) \xrightarrow{p} \ln(1-\tau) \tag{1.16}$$

In contrast, when the baseline model is extended to allow parents to observe shocks to child leisure preference before making discussion decisions, their optimal choice of performance discussion depends on preference shocks to child leisure, as shown in equation (1.17).[65]

$$S_{i,t} = \mathbf{1}\{\beta^p \Gamma^p_{t+1} \alpha_2 \ln \frac{(\lambda_i \epsilon^\lambda_{i,t}) + \beta^c \Gamma^c_{t+1} \alpha_2}{(\lambda_i \epsilon^\lambda_{i,t})(1-\tau) + \beta^c \Gamma^c_{t+1} \alpha_2} > \xi_{i,t}\} \tag{1.17}$$

It can be shown that $\mathbb{P}(S_{i,t} = 1)$ increases in $\epsilon^\lambda_{i,t}$. Intuitively, when child i receives a positive shock to leisure and thus exerts lower effort, her parents feel it more necessary to have performance discussion to influence child i's effort choice. This finding indicates that

$$\mathbb{E}(\ln \epsilon^\lambda_{i,t} | S_{i,t}) \neq \mathbb{E}(\ln \epsilon^\lambda_{i,t}) = 0.$$

Specifically, positive shocks to leisure are more likely to result in performance discussion $S_{i,t} = 1$.

$$\mathbb{E}(\ln \epsilon^\lambda_{i,t} | S_{i,t} = 1) > \mathbb{E}(\ln \epsilon^\lambda_{i,t}) = 0$$
$$\mathbb{E}(\ln \epsilon^\lambda_{i,t} | S_{i,t} = 0) < \mathbb{E}(\ln \epsilon^\lambda_{i,t}) = 0$$

Therefore, the estimated $\hat{\tau}$ using equation (1.16) is smaller than the "true" τ.

$$\begin{aligned} \ln(1-\hat{\tau}) &= \frac{1}{N^\tau} \sum_{i=1}^{N^\tau} \ln(\frac{\bar{L} - L_{i,t^i_1}}{L_{i,t^i_1}} \frac{L_{i,t^i_2}}{\bar{L} - L_{i,t^i_2}}) \\ &\xrightarrow{p} \ln(1-\tau) + \underbrace{\mathbb{E}(\ln \epsilon^\lambda_{i,t^i_1} | S_{i,t^i_1} = 1) - \mathbb{E}(\ln \epsilon^\lambda_{i,t^i_2} | S_{i,t^i_2} = 0)}_{>0} \end{aligned}$$

Estimation Accounts for Reverse Causality

[65] This model extension only affects how parents make discussion decisions. Other choices are the same as the baseline model.

Table 1.13: Bounding Model Parameters: Baseline and Extended Models

	Parameter	Baseline		Extended
disutility from interaction of discussion and leisure	τ	0.549	[0.503, 0.590]	0.716
child's preference for leisure	mean of λ_i	0.207	[0.105, 0.341]	0.487
	sd of λ_i	0.233	[0.115, 0.389]	0.550
parents' preference for consumption	mean of δ_i	1.545	[0.476, 2.506]	1.569
	sd of δ_i	0.943	[0.285, 1.529]	0.957
discussion cost parameters				
mean cost: intercept	π_0	-2.973	[-3.682, -2.460]	-2.160
mean cost: slope of consumption preference	π_1	0.073	[0.028, 0.205]	0.046
mean cost: slope of college-educated mother	π_2	-0.342	[-0.422, -0.280]	-0.222
mean cost: slope of first-born	π_3	-0.150	[-0.237, -0.069]	-0.107
mean cost: slope of girl	π_4	0.118	[0.041, 0.196]	0.064
standard deviation of cost	σ^ξ	0.545	[0.495, 0.616]	0.204
parents' terminal value	ψ^p	1.037	[0.858, 1.258]	1.102
child's terminal value	ψ^c	1.045	[0.847, 1.350]	1.212

Notes: 95% confidence intervals are computed using bootstrap sampling household with replacement.
Production parameters are estimated in the same way in the extended model as the baseline, so they are omitted.

When the assumption that parents do not observe preference shocks to child leisure is relaxed, the estimation strategy in equation (1.16) fails to address the bias from reverse causality. I employ an alternative estimation strategy that accounts for reverse causality in the extended model. The key is to simulate parents' choices of performance discussion (dependent on leisure shocks) and child effort (dependent on discussion choices) according to the extended model.[66] Then, I use indirect inference to estimate τ. I run a regression of effort on performance discussion (with individual fixed effects) using the simulated data and the survey data, respectively. I find the value of τ that predicts the same effect of performance discussion on effort with the simulated data[67] as what is obtained with the survey data.

1.7.2 Bounding Estimation Results

The baseline model ignores reverse causality from effort shocks to performance discussion and thus underestimates the effect of discussion on effort (shown in section 1.7.1). The extended model assumes that parents have full knowledge of effort shocks, which over-corrects the bias from reverse causality if parents actually have partial knowledge of effort shocks in reality. Therefore, the true value of τ is bounded by the baseline estimate and the alternative estimate. Table 1.13 compares the parameter estimates in the baseline and the extended model. The estimated value of τ in the extended model is 0.716, larger than the baseline estimate of 0.549.

The change in the estimated parameter translates into changes in the quantitative results. First, I quantify the contributions of performance discussion to child effort and skill development with the extended model. As shown in table 1.14, the true contributions are bounded by the baseline results and the alternative results. Shutting down the channel of performance discussion would decrease skill accumulation over four years by more than 0.090 standard deviation and less than 0.123 standard deviation. Second, I estimate the impact of eliminating difference in discussion cost across socioeconomic groups on the skill

[66] Leisure shocks directly affect the probability of having performance discussion. I simulate R paths of performance discussion choices (binary) to reflect the change in probability associated with leisure shocks.

[67] The predicted effect is the average of R simulated paths.

Table 1.14: Bounding Results: Contributions of Performance Discussion to Average Skill Development

Shutting down discussion channel	Baseline	Extended
average child effort	↓ 10.55-12.32%	↓ 13.57-16.01%
skill accumulation in four years	↓ 0.090 std. dev.	↓ 0.123 std. dev.

Notes: The Baseline column reports the change in average child effort and four-year skill accumulation when shutting down the channel of performance discussion in the baseline model.
The Extended column reports the change in average child effort and four-year skill accumulation when shutting down the channel of performance discussion in the extended model.

gap with the extended model. As shown in table 1.15, closing the discussion gap would slow down the expansion of skill gap over four years by more than 0.018 standard deviation and less than 0.021 standard deviation.

Table 1.15: Bounding Results: Contributions of Performance Discussion Gap to Skill Gap

Eliminating difference in discussion cost by SES	Baseline	Extended
four-year expansion of skill gap	↓ 0.018 std. dev.	↓ 0.021 std. dev.
equiv. to parental investment of low-SES	↑ \$1,491	↑ \$1,736
equiv. to family income of low-SES	↑ \$19,431	↑ \$22,114

Notes: High-SES: mother is college-educated
The Baseline column reports the change in skill gap when eliminating difference in discussion cost by SES in the baseline model.
The Extended column reports the change in skill gap when eliminating difference in discussion cost by SES in the extended model.

1.8 Conclusion

This paper acknowledges the active role of children in their own skill development, sheds light on multiple channels through which parents influence children's development, and emphasizes the importance of parent-child strategic interaction through performance discussion in addition to parental investment in education. By incorporating parent-child strategic interaction into the framework of dynamic skill accumulation, this paper utilizes the unique KELS-13 panel data on educational inputs and intra-household interactions to estimate a dynamic model that characterizes the two channels of parental influence on child development. In line with previous studies, the estimated model confirms a significant impact of parental monetary investment on human capital accumulation. Moreover, it contributes to the existing literature by establishing an innovative and substantive link from performance discussion to child effort and skill development.

In the context of skill development, the focal point is not limited to improving the average skill. A more challenging goal is to reduce the skill gap between advantaged and disadvantaged groups. Based on the model estimates, this paper finds that heterogeneous parental actions account for a considerable portion of the skill gap between children from high- and low-socioeconomic groups. Furthermore, counterfactual analysis is conducted to unravel the mechanisms underlying disparities in parental actions by leveraging the rich heterogeneity in endowments, preferences, and resources within the model. The main finding is that while divergence in parental educational investment mainly reflects family income inequality, differential performance discussion decisions are largely attributed to heterogeneous household preferences. Identifying the primary factors that drive parental behaviors is vital for making informed policy interventions to promote equitable skill development across diverse backgrounds.

Chapter 2

Closing Migrant-Local Skill Gap: The Roles of School Quality and Parental Investment

2.1 Introduction

Existing literature has documented a significant cognitive skill gap between migrant students and local students in urban areas in China (Lu and Zhou, 2013; Chen and Feng, 2019). One major driver of the gap proposed by the existing studies is the difference in school quality: compared to local students, migrant children face more restrictions on school choices; thus, they tend to enroll in schools with lower quality. However, these studies shed little light on other factors that contribute to the migrant-local skill gap. To help fill this gap, this paper explores another potentially influential factor - parental influence, because home environment and parental investment are shown to have significant impact on child development Heckman and Mosso (2014). In particular, I focus on two types of parental investments: one is parenting activities involving direct parent-child interactions that capture parental non-monetary investment and the other is private tutoring that reflects parental monetary investment. By including parental investments, this paper attempts to quantify and compare the impact from school and family sides on the migrant-local skill gap.

Using nationally representative data from the China Education Panel Survey (CEPS), this paper documents significant gaps in parental investments between rural-to-urban migrant and urban local students in middle school. In terms of parental investment in parenting activities, as shown in table 2.1, migrant children have less educated parents, who are less likely to engage in parenting activities related to cognitive skill development (e.g., helping with homework). As for parental investment in tutoring, the probability of migrant students taking private tutoring is 73 percent, which is seven percentage points lower than that of local students. Conditional on taking tutoring, migrant students spend 1.4 less hours per week (or 12 percent less time) than their local peers. In addition, consistent with existing literature, I find a significant gap in cognitive test scores between migrant and local students: the average cognitive test score of migrant children is 0.08 standard deviation lower than their local counterparts in eighth grade.

To assess the contributions of school quality and parental investments to the skill gap between migrant and local students, I proceed in two steps. First, following the framework in Agostinelli et al. (2025), I build a skill production function that takes into account existing skill, school quality, and parental investments in private tutoring and parenting activities as inputs to produce future cognitive skill. I employ a latent factor model, where all the inputs are latent variables, and parental investments are measured with measurement errors. To examine heterogeneous effects of inputs on cognitive skill formation, I consider the interaction terms of current cognitive skill and educational inputs from schools and parents.

The skill technology is estimated using CEPS data, which provides rich information on students' cognitive skill, parental investments, and school inputs. I find that both school quality and parental investment in parenting activities are important productive inputs in students' cognitive skill development, whereas parental investment in private tutoring is not productive. In terms of average effects, a one standard deviation increase in school quality increases students' cognitive skills in the second period by 0.364 standard deviation, with all other inputs being equal. A one standard deviation increase in parenting activities raises next-year cognitive skill by 0.074 standard deviation, which is 20 percent of the school effects estimates. If ignoring parental investments, the estimated effect of school quality is upward biased, and vice versa, indicating that there exists sorting of students with higher parental investment into better quality schools. Thus, it is necessary to include both parental and school investments in the skill technology.

Another finding is the heterogeneous effects of school and parental investments on skill development. In particular, I find negative complementarity between initial cognitive skill and both school input and parental investment in parenting activities: both investments are more productive for students with lower initial cognitive skills. Specifically, the productivity of school quality (parenting activities) for a student whose initial skill level is one standard deviation below average is 2 (2.75) times that for a student with initial skill one standard deviation above average. This result is in contrast to the typical positive complementarity assumption[1], where investments are more productive for students with higher skills. This finding highlights the importance of adopting a flexible functional form that allows for unrestricted complementarity.

In addition, from the estimated joint distribution of the educational inputs in the skill technology, I find significantly positive correlations among initial skills, school quality, and parental investments in parenting activities and private tutoring. These positive correlations together with the skill technology suggest a persistence in skill development: high-skill students tend to enroll in schools with better quality and receive more parental investments, resulting in higher future skills. In the context of migrant and local students, this means that migrant children, who start with lower skills, face two disadvantages in educational opportunities. First, I find that the average school quality for migrant children is 0.09 standard deviation lower than that of local children. Second, the average investment in parenting activities received by migrant children is 0.54 standard deviation lower than

[1]The commonly used production function with constant elasticity of substitution is one example of positive complementarity assumption.

that of local children.

In the second step, to show how addressing the gaps in educational inputs can reduce the cognitive skill gap between migrant and local students in urban areas, I conduct two counterfactual exercises where migrant students 1) have the same access to schools and 2) receive the same amount of parental investments as local students. The predicted changes in the skill gap in the two exercises reflect the contributions of school quality and parental investments to the skill gap, respectively. First, I find that if migrant children have the same access to schools as local students, the migrant-local skill gap is reduced by 42 percent. Second, providing the average parental investment in parenting activities received by local students to migrant children closes the skill gap by 17 percent. While the relative productivity of parenting activities is much lower than school quality,[2] the contribution of differences in parenting activities to the migrant-local skill gap is more than 40 percent of the contribution of differences in school quality. This is because the school quality difference is much smaller than the parenting activities difference between migrant and local students. This finding emphasizes the relevance of including parental influence in the analysis of the migrant-local skill gap.

The paper is organized in the following way. Section 2 introduces the institutional background of the study and reviews the related literature. Section 3 presents the model. Section 4 describes the data and discusses identification and estimation of the model. Section 5 reports the estimation results and the decomposition exercises. Section 6 conducts several robustness checks for the estimation. Section 7 concludes.

[2] As mentioned above, for a student with the average initial skill, the productivity of parenting activities is 20 percent of that of school quality.

2.2 Institutional Background and Literature

2.2.1 Rural-to-Urban Migration, *Hukou*, and Access to Public Education

Since the economic reform in the 1980s, China has experienced an unprecedented surge of rural-to-urban migration along with rapid economic growth. A great number of rural residents have migrated to the cities in search of more work opportunities (Démurger et al., 2009). However, the majority of them still have their official household registration status, *hukou*, in the rural area. According to the Seventh National Census data, such migrant population without local *hukou* has amounted to 380 million in 2020, which accounts for more than a quarter of the national population (Ning, 2021). Among the non-local *hukou* migrants, approximately 14 million are migrant children in the compulsory education period (Grade 1 to Grade 9).

In China, *hukou* is tied to many public services and social benefits, including health care, pension, and public education. The absence of local *hukou* not only limits the migrant workers' access to urban public services, but also restricts their children's access to public education. Free public compulsory education is provided by the local government at the county level and the funding is allocated by the number of children with local *hukou* and is not portable across counties. Public schools rely on two criteria to admit students. First, students must reside within the local school district. Second, students must be registered in the school district, i.e., having local *hukou*. Therefore, the enrollment in the public schools is guaranteed only for children with local *hukou*.

For a long time, migrant children were admitted to the public schools only if there were extra seats after all local-*hukou* children were enrolled and they were charged high out-of-district tuition fees. In the 1990s, most migrant children were not able to enroll in local public schools in the cities and had to go to the privately operated migrant schools. These private migrant schools had teachers with less experience and worse educational background and facilities of much worse quality compared to public schools (Chen and

Feng, 2013).

Starting in 2008, the central government has implemented a series of inclusive education policies to promote the access to the public schools for migrant children. In 2008, the State Council promised to make bonus transfers to the local provincial governments if they agreed to accommodate more migrant children in their public education system. After a reform of the household registration system in 2014, which substitutes *hukou* with residence permit [3] for migrants when determining the access to public services, the State Council relaxed the tie between *hukou* and public education and enacted a new policy that requires the local government to provide the migrant children the same access to free public schools as their local-*hukou* counterparts in 2016. According to the Ministry of Education, 80 percent of the migrant students in primary and lower secondary education period are enrolled in the public schools in 2020. Nonetheless, anecdotes suggest that migrant children are less likely to enter the top quality public schools than local students.

2.2.2 Private Tutoring: A New Source of Educational Inequality?

Private tutoring has expanded quickly and is becoming a major educational investment besides regular schooling from parents in China in recent decades. According to a nationally representative data source, China Education Panel Survey (CEPS), 72 percent of Chinese students in junior high school report to have received private tutoring in the previous year in 2014.[4] Market-oriented provision of private tutoring results in unequal investment in private tutoring, which could potentially dampen the efforts of government in promoting equal access to public educational opportunities among groups from different socioeconomic background (Bhorkar and Bray, 2018).

To deal with the persistent concerns about private tutoring, in July 2021, the Ministry of Education in China issued a policy that prohibits for-profit private tutoring institutions from providing academic tutoring service to primary and junior high school students nationwide, which aims to reduce the study burden of students in lower grades and to reduce

[3] Migrants with residence permit card can have equal access to public service as residents with local *hukou*, while keeping their *hukou* in their hometown.

[4] See in table B.1 in the appendix.

educational inequality. It is likely that the policy can lower the amount of private tutoring taken by the targeted students, but whether the goal of reducing educational inequality can be achieved or not remains unclear. To predict the impact of reducing private tutoring on educational inequality, the first step is to examine the relationship between the investment in private tutoring and student's cognitive skill. While the amount of private tutoring received by students is substantial, there is no consensus on the effectiveness of private tutoring in the existing literature.[5] Thus, focusing on students in junior high school, the period targeted by the national policy, this paper attempts to estimate the impact of private tutoring on their cognitive skill formation, which will help predict the policy implications of prohibiting private tutoring.

2.2.3 Related Literature

The educational challenges faced by migrant children in China have attracted much attention from researchers in the field of education (Lu, 2007; Chen and Liang, 2012; Goodburn, 2009). One major concern is the observed academic performance gap between migrant and local urban students. Lu and Zhou (2013) is one of the first quantitative studies that attempt to identify the mechanisms behind the migrant-local skill gap. Focusing on primary schools in Beijing, they show that migrant students have lower test scores than local students and that the skill gap is closed by around 15 percent, when moving migrant students from low-quality private migrant schools to high-quality public schools. Chen and Feng (2019) study a more recent sample in Shanghai and focus on the comparison between migrant and local students within the public education system. Analyzing students' school choices, they reach a similar conclusion that the sorting of students to public schools of different quality according to socioeconomic status is an important channel, through which the migrant-local skill gap is formed. Specifically, they find that

[5] Ryu and Kang (2013) use the panel data from the Korean Education Longitudinal Study and propose birth order as an instrumental variable to identify the effect of private tutoring on test scores. They find no evidence for the effectiveness of private tutoring for the junior high school students in urban South Korea. Zhang (2013) uses peers' participation in private tutoring as an instrumental variable and finds a statistically and substantively significant positive effect on the college entrance exam grades of lower-achieving students in urban China.

school quality accounts for around 60 percent of the migrant-local skill gap.

While the difference in school quality is an important factor in explaining the migrant-local skill gap, formal schooling is not the only educational investment that affects academic performance. In particular, studies have shown that home environment and parental investments are shown to have significant impact on skill development Heckman and Mosso (2014). To the best of my knowledge, this is the first study that explores the role of parental investments in the skill gap between migrant students and local students in the urban areas in China.

The methodology of this research is in the same spirit of Agostinelli et al. (2025) with a unified empirical framework that incorporates the key features in the child development literature and the education production function literature. The child development literature (Cunha et al., 2010; Agostinelli and Wiswall, 2025) analyzes the effect of parental investment on child's cognitive skill development, while the education production function literature (Krueger, 1999; Todd and Wolpin, 2003; Chetty et al., 2014) focuses on the influence of school environment on students' academic performance. In the child development literature, complementarities between existing skills and investments are allowed, suggesting heterogeneous effects of investments on skill formation. In the production function literature, the school influences are treated as a latent fixed effect, which is identified when multiple students are observed in the same school. I combine the conventions in both streams of literature. In the unified model that considers investments from both parents and schools, I can decompose the migrant-local skill gap into the influences from parental investments and school quality.

2.3 Model

In this section, I describe the model of cognitive skill development, which follows the general framework of child development in Agostinelli et al. (2025). I discuss the identification of the model and close the section by developing a practical estimator that can be taken to the data.

2.3.1 Cognitive Skill Development

Consider a two-period model for student cognitive skill development. There is a population of students and each student in the population is indexed i. In period 0, each child is characterized by an initial skill stock $\theta_{i,0}$, a flow investment from parents in private tutoring $T_{i,0}$, a flow investment from parents in parenting activities $P_{i,0}$, and a flow investment from school, $S_{i,0}$. In this paper, I consider a scalar cognitive skill and three scalar investments from parents and school. For each student, the initial stock of cognitive skill and initial flows of investments from parents and school produce the stock of cognitive skill in the next period according to the skill formation technology:

$$\theta_{i,1} = h(\theta_{i,0}, T_{i,0}, P_{i,0}, S_{i,0}; \eta_{i,0}) \tag{2.1}$$

where $\eta_{i,0}$ is the unobserved production shock.

The investments and skill stocks are all strictly positive. The investment from school represents all cognitive skill development activities during the school day, including interactions from teachers and peers.

In the remainder of the paper, I follow the functional form of skill technology in Agostinelli and Wiswall (2025):

$$\ln \theta_{i,1} = \ln A + \psi \ln f(\theta_{i,0}, T_{i,0}, P_{i,0}, S_{i,0}) + \eta_{i,0} \tag{2.2}$$

where $f(\theta_{i,0}, T_{i,0}, P_{i,0}, S_{i,0})$ represents a production function whose location and scale (in logarithm) are fully captured by the total factor productivity (TFP) term $\ln A$ and the return to scale parameter ψ; the production shock $\eta_{i,0}$ satisfies the mean-independence assumption:

$$\mathbb{E}(\eta_{i,0} | \theta_{i,0}, T_{i,0}, P_{i,0}, S_{i,0}) = \mathbb{E}(\eta_{i,0}) = 0.$$

where the second equality comes with the normalization given the location term $\ln A$ in the technology (2.2).

In this paper, I would like to employ the production technology to address several quantitative questions. First, it is important to estimate and compare the productivity of various investments. Second, is there any complementarity in the production technology between current skill stocks and investments? In other words, how does the heterogeneity in students' current cognitive skills affect the productivity of investments?

2.3.2 Parametric Specification

The functional form described in this section is useful to answer the questions discussed above. In order to allow for both positive and negative complementarities of initial cognitive skill and investments from parents and school, I specify the following cognitive skill production function similar to Agostinelli et al. (2025):

$$\begin{aligned}\ln \theta_{i,1} = \ln A + \gamma_1 \ln \theta_{i,0} + \gamma_2 \ln T_{i,0} + \gamma_3 \ln P_{i,0} + \gamma_4 \ln S_{i,0} \\ + \gamma_5 \ln \theta_{i,0} \times \ln T_{i,0} + \gamma_6 \ln \theta_{i,0} \times \ln P_{i,0} \\ + \gamma_7 \ln \theta_{i,0} \times \ln S_{i,0} + \eta_{i,0}\end{aligned} \tag{2.3}$$

where the TFP term $\ln A$ represents the intercept of the model and the $\{\gamma_k\}_{k=1}^{7}$ are primitive productivity parameters. In the empirical analysis, $\ln A$ is a function of the student's observable characteristics, such as age, gender and number of siblings. The return of school investment is heterogeneous and depends on the student i's current cognitive skills:

$$\frac{\partial \ln \theta_{i,1}}{\partial \ln S_{i,0}} = \gamma_4 + \gamma_7 \ln \theta_{i,0} \tag{2.4}$$

where the sign of γ_7 indicates whether current skill and school investment are positively or negatively complementary. If $\gamma_7 < 0$, it indicates that the return to school investment for students with lower current skills is higher. Similarly, the return of parental investments is heterogeneous and depends on student i's current skills:

$$\frac{\partial \ln \theta_{i,1}}{\partial \ln T_{i,0}} = \gamma_2 + \gamma_5 \ln \theta_{i,0} \tag{2.5}$$

$$\frac{\partial \ln \theta_{i,1}}{\partial \ln P_{i,0}} = \gamma_3 + \gamma_6 \ln \theta_{i,0} \tag{2.6}$$

2.3.3 Measurement

The cognitive skill production function (2.1) is written in terms of latent variables. Students' skills and the investments from parents and schools are unobserved. Following the previous literature in child skill development, I consider a log-linear system of measures for the latent skill stocks and parental investments.

For each period t, the measurement model for latent skills $\ln \theta_t$ is given by:

$$Z_{\theta,t} = \mu_{\theta,t} + \lambda_{\theta,t} \ln \theta_t \quad \text{for } t = 0, 1$$

where $Z_{\theta,t}$ is the observed measure for $\ln \theta_t$; $\mu_{\theta,t}$ and $\lambda_{\theta,t}$ are the measurement parameters. Due to the data constraint,[6] the cognitive skills are assumed to be perfectly measured, i.e., no measurement errors, which is more restrictive than recent child development literature.

In contrast, since there are multiple measures for parental investments in private tutoring and parenting activities, measurement errors are considered in the measurement model for $\ln T_0$ and $\ln P_0$:

$$Z_{T,m,0} = \mu_{T,m,0} + \lambda_{T,m,0} \ln T_0 + \epsilon_{T,m,0} \quad \text{for } m = 1, ..., M_T$$

$$Z_{P,m,0} = \mu_{P,m,0} + \lambda_{P,m,0} \ln P_0 + \epsilon_{P,m,0} \quad \text{for } m = 1, ..., M_P$$

where $Z_{m,0}$ is the observed measure; $\mu_{m,0}$ and $\lambda_{m,0}$ are the measurement parameters, representing the location and scale of the measure m; $\epsilon_{m,0}$ is the mean-zero measurement error, $E(\epsilon_{m,0}) = 0$.

In line with the education production literature, the investment from school $\ln S_{i,0}$ is treated as a school specific fixed effect. The idea is that school input $\ln S_{i,0}$ is the same for all the students in a given school and the "return" of $\ln S_{i,0}$ on each student can

[6] The dataset (described below in the data section) includes only one measure of cognitive skill that is comparable across schools in the sample.

vary by existing cognitive skill (captured by the interaction term of $\ln S_{i,0}$ and $\ln \theta_{i,0}$). With multiple students surveyed in each school, the distribution of $\ln S_{i,0}$ is recovered via a generalized non-linear fixed-effect estimator following Arcidiacono et al. (2012), as described in more details in the estimation section.

2.4 Data and Estimation

2.4.1 Sample Construction

This study employs a nationally representative panel dataset from the China Education Panel Survey (CEPS). Based on the average educational level and the proportion of migrating population, the CEPS data is a stratified sample of approximately 20,000 junior high school students in two grades in 438 classrooms from 112 schools across 28 counties in mainland China. The 2013 baseline survey of students in Grade 7 and Grade 9 includes 5 different questionnaires for students, parents, subject teachers, head teachers, and school principals. All Grade 7 students in the baseline survey were supposed to be interviewed again in a follow-up survey when they were in Grade 8. In practice, 91.9 percent of them were followed successfully. The 830 missing cases are mainly due to transferring to other schools (71%), drop-out (15%), and failure to show-up during the interview date (10%).

In order to be consistent with the two-period production function model, I restrict the sample to include only the group of students who are surveyed twice. In addition, aiming to investigate the skill gap between migrant and local students in the urban area, I exclude 3,920 students who are rural non-migrant students. Furthermore, I drop 283 observations from the non-public schools, because unlike public schools, non-public schools are not free and the tuition fees are another type of parents' monetary investment in education, which makes their educational investment decisions different from those in the public schools.[7]

[7] As shown in Table B.2 in the Appendix, migrant students in the public schools have similar characteristics as all migrant students in both public and non-public schools in terms of demographics, test scores, parental investments in private tutoring and parenting activities, educational aspirations, and household income level. Parents of migrant students in public schools have 0.2 more years of schooling than those of all migrant students do.

Finally, I keep only observations with non-missing test scores, private tutoring information, and demographic information, including gender, age, family size, and parents' education information, which drops 69 observations. As a result, the baseline sample size is 5,176, of which migrant students account for approximately 30 percent.[8]

2.4.2 Measures of Variables

Cognitive Skill

The CEPS cognitive ability assessment battery attempts to evaluate the cognitive skill development of students over time. The assessments are designed to be comparable across periods in order to track students' learning. The cognitive ability test contains questions from three subject areas: language, geometry, and algebra. The cognitive test score estimated from the three-parameter Item Response Theory (IRT) model is recorded in the dataset.

Private Tutoring

The CEPS dataset reports two measures for parental monetary investment in private tutoring. First, the survey asks for the average daily time spent on private tutoring on both weekdays and weekends separately and a weighted daily average of weekday and weekend is computed. Second, total expenditure on private tutoring in the surveyed semester is reported.

Parenting Activities

In terms of (non-monetary) parenting activities related to general cognitive skill formation, the CEPS dataset contains information on multiple activities including helping

[8] As shown in Table B.1 in the Appendix, students in the baseline sample are different from the full sample in the following ways. First, in terms of demographics, students in the baseline sample are younger, 9 percentage points more likely to be a migrant student, 15 percentage points more likely to be the only child, and from richer families. Second, their parents are more educated and invest more in both private tutoring and parenting activities. Third, they have higher cognitive test scores and they are 5 percentage points more likely to plan to receive college education. In general, the selected sample is more socioeconomically advantaged than the full sample.

with the child's homework, reading with the child, visiting museums with the child, etc. In contrast to monetary investment in tutoring, these parenting activities involve more direct interactions between parents and children, the quality of which depends on parental cognitive skill. Following the idea in Cunha et al. (2010) that finds positive correlations between parental investment in cognitive activities and parents' educational attainment and treats parents' cognitive skill, measured by educational attainment, as a productive input of children's skill formation, I use the parents' years of schooling to measure the quality of parenting activities.[9]

2.4.3 Identification

The model is identified up to some initial normalization given that latent cognitive skills and investments are not directly observed. I normalize all the initial period latent variables to be mean of 0 and variance of 1.

$$E(\ln \theta_{i,0}) = E(\ln T_{i,0}) = E(\ln P_{i,0}) = E(\ln S_{i,0}) = 0$$

$$V(\ln \theta_{i,0}) = V(\ln T_{i,0}) = V(\ln P_{i,0}) = V(\ln S_{i,0}) = 1$$

The normalization treats all latent variables symmetrically and eases the interpretation of estimates. It also resolves the arbitrariness of the measures. Any positive monotonic transformation of the measure is also a valid measure.

The normalization of school investment implies that the school effect is represented by $\gamma_4 \ln S_{i,0}$, which means that the standard deviation of the school effects distribution is equal to the parameter γ_4. Therefore, the parameters in equation (2.3), which represent the productivity of each input, can be easily compared.

[9]The reasons of not using the reported frequency of each parenting activity to measure the quality of parenting activities are as follows. First, the reporting scale and period vary across activities. For example, children are asked 1) how often your parents help with your homework *last week* (choices: never, once or twice a week, three or four times a week, almost everyday), v.s., 2) how often your parents read books with you *in general* (choices: never, once a year, twice a year, once a month, once a week, twice a week or more). Second, even with comparable frequency across activities, the quality of these parenting activities varies with parental cognitive skill.

Given the normalization of the latent skills and parental investments in the initial period, the measurement intercepts, μs , are identified from the mean of the observed measures in period $t = 0$.

$$\mu_{\omega,0} = E(Z_{\omega,i,0}) \text{ for all } \omega \in \{\theta, T, P\}$$

Given the normalization of the variance of the latent skills and no measurement errors, the scaling parameter, $\lambda_{\theta,0}$, is identified from the variance of the observed measure in period $t = 0$.

$$\lambda_{\theta,0} = \sqrt{V(Z_{\theta,i,0})}$$

Given the identification of the measurement parameters for $\ln \theta_0$ in the initial period, the latent cognitive skills are identified in period $t = 0$.

$$\ln \theta_{i,0} = \frac{Z_{\theta,i,0} - \mu_{\theta,0}}{\lambda_{\theta,0}} = \tilde{Z}_{\theta,i,0} \tag{2.7}$$

The scaling parameters, $\lambda_{T,0}$ and $\lambda_{P,0}$, in period $t = 0$ cannot be identified without further restrictions on the measurement errors. The following independence assumptions are commonly used in the literature that the measurement errors are independent of each other and of the latent variables:

1. $\epsilon_{\omega,i,m,0} \perp \epsilon_{\omega,i,m',0}$ for all latent variable ω, i, and $m \neq m'$: measurement errors are independent contemporaneously across measures.

2. $\epsilon_{\omega,i,m,0} \perp \epsilon_{\omega',i,m,0}$ for all m, i, and latent variable $\omega \neq \omega'$: measurement errors are independent contemporaneously across different measures of different factors.

3. $\epsilon_{\omega,i,m,0} \perp \omega'$ for all m, i, and latent variables (ω, ω'): measurement errors are independent of the latent children skills and latent parental investments.

4. $\epsilon_{\omega,i,m,0} \perp \ln S_{i,0}$ for all m, i, and latent variable ω: measurement errors are independent of the latent school input.

Under these assumptions, the initial period ($t = 0$) scaling parameters are identified from ratios of covariances between the measures.[10]

$$\lambda_{\omega,m,0} = \sqrt{\frac{Cov(Z_{\omega,m,0}, Z_{\omega,m',0})Cov(Z_{\omega,m,0}, Z_{\theta,0})}{Cov(Z_{\omega,m',0}, Z_{\theta,0})}} \text{ for all } \omega \in \{T, P\} \text{ and } m \neq m'$$

Given the identification of the measurement parameters for $\ln T_{i,0}$ and $\ln P_{i,0}$ in the initial period, the latent investments are identified up to the measurement errors.

$$\ln \omega_{i,0} = \frac{Z_{\omega,i,m,0} - \mu_{\omega,m,0}}{\lambda_{\omega,m,0}} - \frac{\epsilon_{\omega,i,m,0}}{\lambda_{\omega,m,0}} = \tilde{Z}_{\omega,i,m,0} - \tilde{\epsilon}_{\omega,i,m,0} \text{ for all } \omega \in \{T, P\} \tag{2.8}$$

By substituting the measures for both periods $t = \{0, 1\}$ into equation (2.3), the empirical analog of the cognitive skill production technology is given as:

$$\begin{aligned} Z_{\theta,i,1} = \mu_{\theta,1} &+ \lambda_{\theta,1} \ln A + \lambda_{\theta,1}\gamma_1 \tilde{Z}_{\theta,i,0} + \lambda_{\theta,1}\gamma_2 \tilde{Z}_{T,i,m,0} + \lambda_{\theta,1}\gamma_3 \tilde{Z}_{P,i,m,0} \\ &+ \lambda_{\theta,1}\gamma_4 \ln S_{i,0} + \lambda_{\theta,1}\gamma_5 \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{T,i,m,0} + \lambda_{\theta,1}\gamma_6 \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{P,i,m,0} \\ &+ \lambda_{\theta,1}\gamma_7 \tilde{Z}_{\theta,i,0} \times \ln S_{i,0} + \kappa_{i,m,0} \end{aligned} \tag{2.9}$$

where $\kappa_{i,m,0} = \lambda_{\theta,1}(\eta_{i,0} - \gamma_2\tilde{\epsilon}_{T,i,m,0} - \gamma_3\tilde{\epsilon}_{P,i,m,0} - \gamma_5\tilde{Z}_{\theta,i,m,0}\tilde{\epsilon}_{T,i,m,0} - \gamma_6\tilde{Z}_{\theta,i,m,0}\tilde{\epsilon}_{P,i,m,0})$.

The equation (2.9) can be re-written in "reduced form" as:

$$\begin{aligned} Z_{\theta,i,1} = \beta_0 &+ \beta_1 \tilde{Z}_{\theta,i,0} + \beta_2 \tilde{Z}_{T,i,m,0} + \beta_3 \tilde{Z}_{P,i,m,0} + \beta_4 \ln S_{i,0} + \beta_5 \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{T,i,m,0} \\ &+ \beta_6 \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{P,i,m,0} + \beta_7 \tilde{Z}_{\theta,i,0} \times \ln S_{i,0} + \kappa_{i,m,0} \end{aligned} \tag{2.10}$$

where the set of "reduced-form" parameters $\{\beta_j\}_{j=0}^{7}$ are functions of structural and measurement parameters.

$$\beta_0 = \mu_{\theta,1} + \lambda_{\theta,1} \ln A$$

$$\beta_j = \lambda_{\theta,1}\gamma_j, \text{ for } j = 1, 2, ..., 7$$

[10] At least two measures per latent variable are required to identify λ.

The production function parameters, $\ln A$ and γs, are under-identified unless the measurement parameters, $\mu_{\theta,1}$ and $\lambda_{\theta,1}$, of period 1 skill are known. The problem can be solved with some *age-invariant* measure for skills. Specifically, a pair of measures Z_t and Z_{t+1} for latent variables θ_t and θ_{t+1} is age-invariant if $E(Z_t|\theta_t = p) = E(Z_{t+1}|\theta_{t+1} = p)$ for all $p \in \mathbb{R}_{++}$. This assumption implies that the measurement parameters are constant throughout the periods, i.e., $\mu_{\theta,1} = \mu_{\theta,0}$ and $\lambda_{\theta,1} = \lambda_{\theta,0}$. Since the CEPS cognitive ability assessment is designed to be comparable across periods in order to track students' learning, it is reasonable to assume that the cognitive test score is an age-invariant measure for skills.

Another identification challenge comes from selection bias with respect to the structural production shock η. A bias would exist if any of the inputs in the production function, i.e., students' initial skill, parental investments, and school inputs, are correlated with the unobserved shock. Identification of the model requires the following mean-independence assumption of the production function shock:

$$\mathbb{E}(\eta_{i,0}|\theta_{i,0}, T_{i,0}, P_{i,0}, S_{i,0}) = 0.$$

This assumption is weaker than most child development literature, because the mean-independence holds only within each school, not across schools. In other words, it allows for unrestricted sorting into schools based on child skill and parental investments.

2.4.4 Estimation

The inputs from school $\ln S_{i,0}$ are treated as latent school fixed effects with no explicit measures and are estimated as the average within-school residual in new cognitive skills with data using the model. Since $\ln S_{i,0}$ enters the cognitive skill production function non-linearly (interacted with latent skills), I use the recursive estimation algorithm, proposed in Arcidiacono et al. (2012), to estimate the production function parameters and the school fixed effects simultaneously.

After estimating the initial conditions and measurement parameters, the estimation algorithm for the production function parameters takes two recursive steps. In the first step, given a set of production function parameters, the school latent quality distribution is estimated. In the second step, a new set of production function parameters are estimated given the updated school quality distribution in the first step. I repeat the two steps until the parameter estimates converge to a fixed point. The recursive algorithm is presented below.

Step 1: Given the current parameter guess, $\{\ln A^n, \{\gamma_k^n\}_{k=1}^7\}$, estimate the school fixed effects as the average within-school residual in new cognitive skills:

$$\ln S_{i,0}^n = \frac{\sum_{j \in S(i)} \big(\tilde{Z}_{\theta,j,1} - \ln A^n - \gamma_1^n \tilde{Z}_{\theta,j,0} - \gamma_2^n \tilde{Z}_{T,j,m,0} - \gamma_3^n \tilde{Z}_{P,j,m,0} - \gamma_5^n \tilde{Z}_{\theta,j,0} \times \tilde{Z}_{T,j,m,0} - \gamma_6^n \tilde{Z}_{\theta,j,0} \times \tilde{Z}_{P,j,m,0}\big)}{\sum_{j \in S(i)} (\gamma_4^n + \gamma_7^n \tilde{Z}_{\theta,j,0})}$$

where $S(i)$ is the set of students in the school that student i attends.

Step 2: Given the distribution of school fixed effects $\ln S_{i,0}^n$ from Step 1, estimate the empirical analogue of the cognitive skill production function in equation (2.3):

$$\begin{aligned}
\tilde{Z}_{\theta,i,1} = \ln A^{n+1} &+ \gamma_1^{n+1} \tilde{Z}_{\theta,i,0} + \gamma_2^{n+1} \tilde{Z}_{T,i,m,0} + \gamma_3^{n+1} \tilde{Z}_{P,i,m,0} + \gamma_4^{n+1} \ln S_{i,0}^n \\
&+ \gamma_5^{n+1} \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{T,i,m,0} + \gamma_6^{n+1} \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{P,i,m,0} \\
&+ \gamma_7^{n+1} \tilde{Z}_{\theta,i,0} \times \ln S_{i,0}^n + \kappa_{i,m,0}
\end{aligned}$$

where $\kappa_{i,m,0} = \eta_{i,0} - \gamma_2 \tilde{\epsilon}_{T,i,m,0} - \gamma_3 \tilde{\epsilon}_{P,i,m,0} - \gamma_5 \tilde{Z}_{\theta,i,m,0} \tilde{\epsilon}_{T,i,m,0} - \gamma_6 \tilde{Z}_{\theta,i,m,0} \tilde{\epsilon}_{P,i,m,0}$.

Since $\kappa_{i,m,0}$ contains measurement errors, even with known school inputs, OLS estimation of the Step 2 equation would produce an inconsistent estimate of the structural parameters. Therefore, the equation in Step 2 is estimated via 2SLS estimation using the multiple excluded measures of parental investments $(Z_{T,i,m',0},\, Z_{P,i,m',0})$ for some $m' \neq m$ as instrumental variables. This produces the $n+1$ iteration of parameters, $\{\ln A^{n+1}, \{\gamma_k^{n+1}\}_{k=1}^7\}$, which can be used in Step 1.

The procedure stops when all the parameters converge.

2.5 Results

2.5.1 Summary Statistics

Table 2.1 presents the summary statistics for students in the sample. Migrant students refer to those who do not have local *hukou*. In terms of cognitive test scores, on the one hand, both migrant and local students have higher test scores in the second wave of survey than that in the base year, suggesting that the cognitive test tracks their improvement in performance with additional year of schooling. On the other hand, local students have a statistically and substantively significantly higher average test score than migrant students in both years.

In the sample, local students are slightly younger than migrant students: the average age of the former is 13.44 years and that of the latter is 13.55 years. Local students are 29 percentage points more likely to have no sibling, i.e., to be the only child in family. However, they do not show statistically or substantively significant difference in the self-reported three-category family economic status.

As for parents' investment in private tutoring, migrant children on average are 7 percentage points less likely to participate in private tutoring. Those who take tutoring classes spend 0.20 fewer hours per day on private tutoring than the local children.[11] When conditioning on taking tutoring classes, the monthly expenditures do not differ significantly between migrant and local children. In terms of the quality of parenting activities, the parents' of local students are more educated than those of migrant students. On average, the local mother has 1.83 more years of schooling and the local father has 1.49 more years of schooling. Local students have better relations with both parents than migrant students. Local parents also invest more in parenting activities related to cognitive skill development, including helping with children's homework, reading with children, and taking children to shows and movies. The fraction of local children/parents that have educational aspirations for at least college education is higher than their migrant counterparts.

[11] The distribution of tutoring hours and spending for migrant and non-migrant populations are shown in Figure B.1 in the Appendix.

Table 2.1: Summary Statistics

	(1) Migrant		(2) Local		Diff.
	Mean	SD	Mean	SD	(2)−(1)
Demographics					
Age	13.55	(0.68)	13.44	(0.61)	-0.11***
Fraction of male students	0.52		0.51		-0.01
Fraction of no sibling	0.38		0.67		0.29***
Fraction of one sibling	0.45		0.25		-0.20***
Fraction of two siblings	0.12		0.05		-0.07***
Fraction of three or more	0.04		0.02		-0.02**
Fraction of rich family	0.06		0.07		0.01
Fraction of middle income	0.79		0.80		0.01
Fraction of poor family	0.15		0.13		-0.02
Mother's years of schooling	9.54	(3.35)	11.37	(3.54)	1.83***
Father's years of schooling	10.50	(3.04)	11.99	(3.28)	1.49***
Test Scores					
Initial cognitive test score	0.03	(0.86)	0.19	(0.89)	0.16***
New cognitive test score	0.36	(0.81)	0.44	(0.83)	0.08**
Parental Investments					
Fraction taking tutoring	0.73		0.81		0.07***
Conditioning on taking tutoring...					
- Daily tutoring hours	1.47	(1.18)	1.67	(1.29)	0.20***
- Monthly fees (¥1K)	0.68	(0.93)	0.70	(1.03)	0.02
Good relation with mother	0.72		0.77		0.06***
Good relation with father	0.62		0.66		0.05**
Fraction doing the following activities with child					
- Help with Homework	0.64		0.70		0.05***
- Reading	0.67		0.71		0.04**
- Watch TV	0.94		0.93		-0.01
- Play sports	0.72		0.77		0.05***
- Visit museum	0.84		0.85		0.01
- Watch shows	0.65		0.72		0.07***
Educational Aspiration					
Going to college (child)	0.83		0.87		0.04***
Going to college (parents)	0.83		0.86		0.04**
No. of Observations	1,501		3,675		

Source: CEPS

2.5.2 Measurement Parameters

Table 2.2 presents the estimates of measurement parameters for the measures of latent variables: cognitive skills and parental investments in private tutoring and parenting activities. The location and scale of each measure is arbitrary and determined by the initial period normalization of latent variables. As described above, the latent variables in the first period is normalized to be mean 0 and variance 1. Thus, the location of a measure is the mean value of the measure. For the cognitive test measure, a location of 0.16 indicates that students have an average score of 0.16 on the test in the base year of survey. For the perfectly measured cognitive skill, the scale of its measure is the standard deviation of the measure; for the parental investments measured with errors, the scale of a measure is identified from the correlation among the measures. The scale can be interpreted as the effect of one standard deviation change in the latent variable: an increase of one standard deviation in the latent skill means an average increase of 1.01 units in the cognitive test score.

Table 2.2 also reports the signal-to-noise ratio for each measure at the initial period. The signal-to-noise ratio is the fraction of the variance of the measure that is explained by the latent variable.[12] The signal-to-noise ratio is identified from the correlation of the measures: a higher correlation across measures implies lower noise. A higher signal-to-noise ratio indicates that the measure is more informative about the latent variable. For example, the signal-to-noise ratio of 0.70 for father's years of schooling indicates that 70 percent of the measure's variation is due to the latent variable.

[12]The signal-to-noise ratio for the measures of cognitive skill without measurement errors is 1 by construction.

Table 2.2: Estimates of Measurement Parameters at Initial Period

Latent Variable	Measure	μ	λ	Signal-to-Noise Ratio
$\ln \theta_0$	Cognitive test score	0.16	1.01	1.00
		(0.014)	(0.009)	(0.000)
$\ln T_0$	Daily tutoring hours	1.27	0.43	0.11
		(0.017)	(0.083)	(0.041)
	Monthly tutoring spending	0.34	0.69	0.77
	(¥1000)	(0.011)	(0.214)	(1.273)
$\ln P_0$	Father's years of schooling	11.56	2.75	0.70
		(0.044)	(0.079)	(0.037)
	Mother's years of schooling	10.84	2.91	0.66
		(0.051)	(0.085)	(0.036)

Standard errors in parentheses

Source: CEPS

Notes: The estimates are for the initial period (t = 0). For each measure $Z_{\omega,t,m}$ of latent variable ω at time t, the location is $\mu_{\omega,t,m}$, the scale is $\lambda_{\omega,t,m}$, and the signal-to-noise ratio is $1 - \frac{\mathbb{V}(\epsilon_{\omega,t,m})}{\mathbb{V}(M_{\omega,t,m})}$.

Table 2.3: The Cognitive Skill Production Function

$\ln \theta_{i,1}$	(1)	(2)		(3)		(4)	
		OLS	IV	OLS	IV	OLS	IV
$\ln \theta_{i,0}$	0.409***	0.478***	0.463***	0.401***	0.395***	0.422***	0.416***
	(0.0170)	(0.0292)	(0.0294)	(0.0168)	(0.0174)	(0.0145)	(0.0153)
$\ln S_{i,0}$	0.389***			0.383***	0.385***	0.360***	0.364***
	(0.00648)			(0.00746)	(0.00646)	(0.00640)	(0.00596)
$\ln T_{i,0}$		-0.00641	0.0165	-0.0171***	-0.0363***	-0.0183***	-0.0386***
		(0.00467)	(0.0165)	(0.00364)	(0.00476)	(0.00387)	(0.00507)
$\ln P_{i,0}$		0.101***	0.154***	0.0447***	0.0685***	0.0510***	0.0738***
		(0.0141)	(0.0301)	(0.00942)	(0.0130)	(0.00990)	(0.0135)
$\ln \theta_{i,0} \times \ln S_{i,0}$						-0.115***	-0.116***
						(0.0140)	(0.0154)
$\ln \theta_{i,0} \times \ln T_{i,0}$						0.00330	0.00669
						(0.00382)	(0.0102)
$\ln \theta_{i,0} \times \ln P_{i,0}$						-0.0366***	-0.0346*
						(0.0110)	(0.0179)
No. of Students	5,176	5,176	5,176	5,176	5,176	5,176	5,176
No. of Schools	104	104	104	104	104	104	104

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

Source: CEPS

Notes: All models control for students' gender, age, and number of siblings. IV estimates are corrected for measurement errors and OLS estimates are not.

2.5.3 Production Function

Table 2.3 presents the estimates of the production functions that have different specifications. Model (1) is the simplest model, where only the initial cognitive skill and school investment are included as inputs of the production function and no interactions of current skill and investments are considered. This specification is analogous to the model in Chen and Feng (2019) where the focus is how school quality affects the migrant-local skill gap. Model (2) is similar to model (1) except that it is parental investments rather than school investments that are included as inputs. Model (3) includes both parental and school investments, but no interactions of current skill and investments are considered. Model (4) reports the estimates of the preferred specification in equation (2.3), where heterogeneous effects of investments are allowed. Models with imperfectly measured parental investments are estimated with both the OLS estimator where measurement errors are not corrected and the IV estimator where measurement errors are corrected.

Prior Skills

Model (1) in Table 2.3 indicates that initial cognitive skill is important in the production of the skill in the next period. Given the log-log form of the production function, the estimate can be interpreted as an elasticity: a one percent increase in latent cognitive skills in the base year leads to a 0.409 percent increase in cognitive skills in the second year. As shown in other models, adding school and/or parental inputs does not affect the productivity of or elasticity in prior skills significantly.

School Effects

In addition to initial cognitive skills, models (1) and (3) in Table 2.3 also include a school input, treated as a school specific fixed effect. As discussed above, the school investments are normalized to be mean 0 and standard deviation 1, as with all of the latent variables in the first period. Therefore, there is a free parameter to be estimated on the school effect γ_4, and given the normalization, this parameter can be interpreted as the standard deviation of the school effects: $\sqrt{V(\gamma_4 \ln S_{i,0})} = \gamma_4$.

The results for models (1) and (3) indicate that school inputs have sizable effects on the production of cognitive skills, regardless of including parental investments or not. For example, the elasticity of new cognitive skills with respect to school investments in model (1) is 0.389, which is 95 percent the corresponding elasticity of new cognitive skills with respect to initial cognitive skills. This suggests that higher school investment can remediate a substantial proportion of the existing skill gap.

Private Tutoring

Model (2) in Table 2.3 shows that when ignoring school inputs, parental investments in private tutoring has a statistically insignificant effect on cognitive skill development, no matter whether the measurement error is corrected or not. In contrast, when including both parental and school inputs as in model (3), the investment in private tutoring is statistically significantly counter-productive in building the general cognitive skill.[13] With measurement errors corrected, the magnitude of the coefficient doubles, but still, the magnitude of the coefficient is rather small, only about one tenth of the effect of school inputs.

The negative effect of private tutoring on cognitive test score is somewhat surprising. Therefore, I conduct additional analysis to investigate the effect of tutoring. Table 2.4 presents the effect of subject-specific private tutoring on subject test scores. The main explanatory variable is a dummy variable that takes the value of 1 if the student takes private tutoring classes for a given subject. Previous test scores, students' demographics, other parental investments and school fixed effects are controlled. All test scores are standardized to be mean 0 and variance 1. I find statistically significantly positive effect of subject-specific tutoring on the given subject test scores. For example, by participating in math tutoring classes, a student's math test score increases by 0.07 units. One caveat is that the subject test scores are from school-specific exams and are not comparable across

[13] Table B.3 examines how the effect of private tutoring on cognitive skill fades away (and even becomes negative) when gradually adding other inputs/controls (existing skill, demographics, parenting activities, and school fixed effects).

schools, so the previous finding is based on a single school.[14] Nevertheless, the result provides some evidence that private tutoring is effective in increasing subject test scores.

Table 2.4: The Effect of Private Tutoring on Subject Test Scores

Subject-specific test scores	(1) Math	(2) Chinese	(3) English
Tutoring for Math	0.0668***		
	(0.0242)		
Tutoring for Chinese		0.1000***	
		(0.0252)	
Tutoring for English			0.0518***
			(0.0191)
No. of Students	137	137	137

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

Source: CEPS
Notes: The subject-specific test scores are from midterms held by each school and thus are not comparable across schools. The main explanatory variable is a dummy variable that takes the value of 1 if the student takes private tutoring classes for a given subject. Previous test scores, students' demographics, other parental investments and school fixed effects are controlled. The results above are for the school with the largest number of students in the baseline sample. The results are similar when using other three schools with more than 100 students.

The previous finding that tutoring is not productive in promoting general cognitive test scores could be due to the possibility that the exercises students do in tutoring classes focus on improving their test-taking skills and mastering of specific textbook materials, rather than promoting general cognitive skills. Since the cognitive test in the CEPS survey is designed to be irrelevant of specific textbook contents and to have a different format from the subject-specific exams, it is possible that private tutoring is not effective in increasing cognitive test scores. Furthermore, the counter-productivity of tutoring on cognitive skills may suggest that if students are trained to rely heavily on test-taking skills and specific textbook materials to get good grades in exams, their development of general cognitive

[14]The coefficients vary slightly when using different schools, but the sign remains positive.

skill may be hindered.

In Table 2.5, by adding a squared term of tutoring investments to model (3) of Table 2.3, I investigate the potential heterogeneous effect of tutoring investments by intensity of tutoring for subgroups with different initial cognitive skills. Decreasing returns of tutoring investment are found for students with initial cognitive scores lower than the 75^{th} percentile.

Table 2.5: The Non-linear Effect of Private Tutoring by Subgroups

$\ln \theta_{i,1}$	(1) 1^{st} quartile	(2) 2^{nd} quartile	(3) 3^{rd} quartile	(4) 4^{th} quartile
$\ln T_{i,0}$	0.0284	0.0359	0.0295	-0.0363
	(0.0488)	(0.0400)	(0.0361)	(0.0292)
$(\ln T_{i,0})^2$	-0.0162*	-0.0195**	-0.0160**	0.00220
	(0.00921)	(0.00771)	(0.00699)	(0.00556)
No. of Students	1,290	1,293	1,287	1,293

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

Source: CEPS
Notes: The baseline sample are divided into four subgroups according to $\ln \theta_{i,0}$. For example, the first quartile refers to students with initial cognitive scores lower than the 25^{th} percentile. A squared term of tutoring investments $(\ln T_{i,0})^2$ is added to model (3) of Table 2.3.

Parenting Activities

Models (2) and (3) in Table 2.3 show that no matter whether school inputs are included or not, parental investment in parenting activities have a statistically and substantively significantly positive effect on cognitive skill development. However, the magnitude of the effect shrinks by around 55 percent when school inputs are included, suggesting that there exists sorting of the students with higher parental investments to schools with better quality. With measurement error corrected, the magnitude of the coefficient increases by around 50 percent. The IV estimates of model (3) shows that the elasticity of cognitive skills in the second year with respect to parental investments in parenting activities is 0.07, which is approximately 18 percent of the elasticity with respect to school investments, indicating that the productivity of school inputs is relatively higher than that of parenting activities.

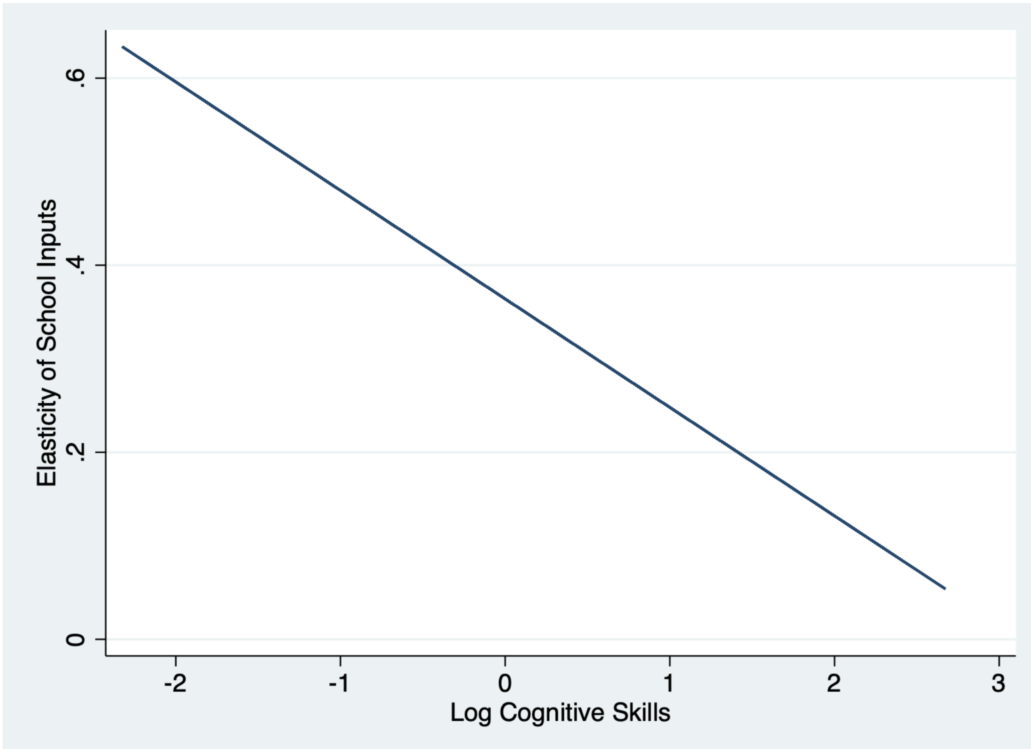
Heterogeneous Effects

Model (4) in Table 2.3 shows the estimates of equation (2.3), which includes potential complementarity between the student's current stock of skills and investments from both school and parents. Complementarity in the model implies heterogeneity in the effect of both school and parental investments: depending on the initial stock of cognitive skill, students will receive different "returns" from a given quality of school or a given amount of parental investments.

The results in Model (4) of Table 2.3 documents a negative complementarity between initial cognitive skills and school inputs, indicating that school investments are more productive for students with low initial cognitive skill relative to students with high initial skill. The complementarity between initial cognitive skill and parental investment in parenting activities is negative as well, indicating that the investments in parenting activities are also more productive for students with low initial cognitive skill than those with high initial skill. Both of the interaction terms discussed above are statistically different from zero at a significance level of 0.1. These results are in contrast to the typical positive complementarity assumption, where investments are more productive for students with higher skills and thus reinforce the skill gap. The negative complementarity implies that some policy interventions can be effective in reducing the skill gaps among students, which will be discussed in more details with the decomposition exercise in the following section. As for the parental investment in private tutoring, the interaction term is not statistically different from zero at standard levels, failing to reject homogeneous effect of private tutoring for students with different levels of cognitive skills.

To better interpret the heterogeneity in the school effects on students, I graph the implied elasticity of school investments with respect to the initial stock of a student's

Figure 2.1: Estimates of Elasticity with Respect to School Inputs

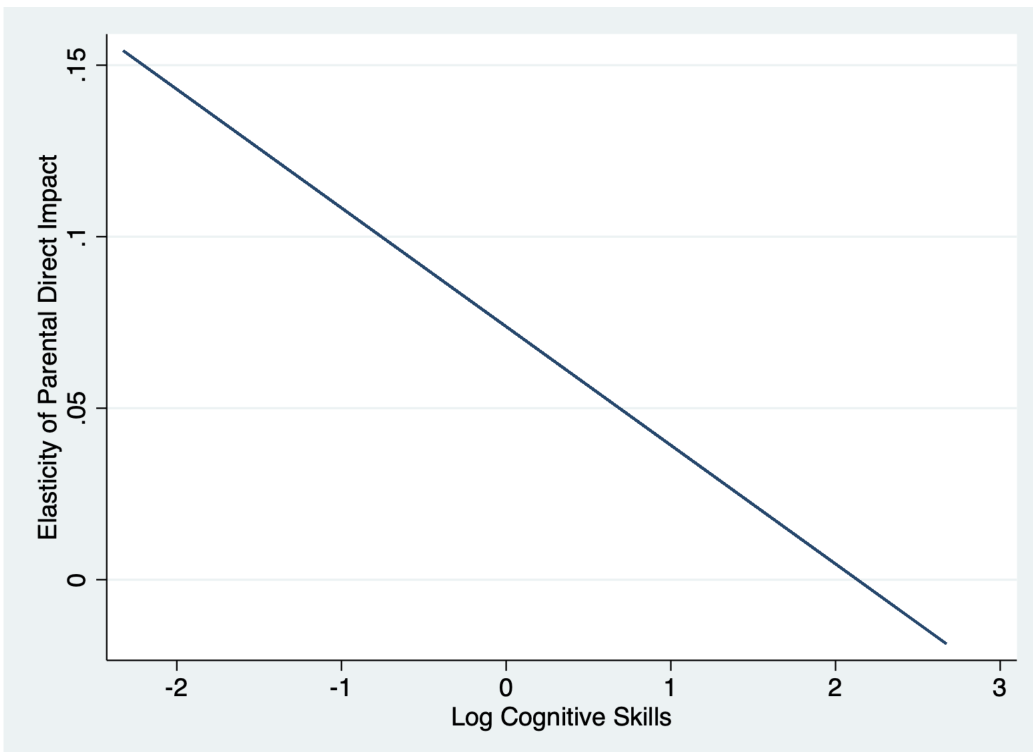


Source: CEPS
Notes: This figure shows the estimates of the elasticity of students' cognitive skills in the second year with respect to school inputs: $\frac{\partial \ln \theta_{i,1}}{\partial \ln S_{i,0}} = \gamma_4 + \gamma_7 \ln \theta_{i,0}$. Results pertain to the model with complementarities and measurement errors corrected (see Model (4) - IV in Table 2.3).

cognitive skills in Figure 2.1. The estimated negative coefficient on the interaction term indicates that the elasticity of cognitive skill production with respect to school investment is decreasing in the student's current cognitive skill. A one standard deviation increase in school investments increases the new cognitive skill of a student with initial cognitive skill level that is 1 standard deviation below the mean level by approximately 0.50 standard deviations, whereas the same change in the school investments increases the cognitive skill of a student with initial cognitive skill level that is 1 standard deviation above the mean level by only 0.25 standard deviations. The effect of school investments are 2 times higher for the former student with lower skill than for the latter higher-skill student.

The heterogeneity of parenting activities effects with respect to initial cognitive skills displays a similar pattern with a flatter slope in Figure 2.2. A one standard deviation increase in parental investments in parenting activities increases the new cognitive skill of a student with initial cognitive skill level that is 1 standard deviation below the mean level by 0.11 standard deviations. In contrast, the same change in parenting activities increases the cognitive skill of a student with initial cognitive skill level that is 1 standard deviation above the mean level by 0.04 standard deviations instead.

Figure 2.2: Estimates of Elasticity with Respect to Parental Investment in Parenting Activities



Source: CEPS
Notes: This figure shows the estimates of the elasticity of students' cognitive skills in the second year with respect to parents' investment in parenting activities: $\frac{\partial \ln \theta_{i,1}}{\partial \ln P_{i,0}} = \gamma_3 + \gamma_6 \ln \theta_{i,0}$. Results pertain to the model with complementarities and measurement errors corrected (see Model (4) - IV in Table 2.3).

2.5.4 Decomposition

In this section, I use the estimated production function to quantify the contribution of school and parental investments in closing the skill gap between migrant and local students. First, I describe the estimated initial joint distribution of latent cognitive skills, school quality, and parental investments. Second, I specifically investigate the matching of students to schools and parents by migrant status, which is not directly part of the model, but can be strongly related to the distribution of initial conditions. Lastly, I conclude with counterfactual exercises in which I manipulate the initial conditions and quantify how the reallocation affects the level of ex post inequality.

Joint Distribution of Initial Cognitive Skills and Investments

I use the estimated model to describe the "initial conditions" of the skill production function: the joint distribution of latent initial skills, school quality and parental investments. The latent school quality, which has no direct measure, is inferred from the estimated outcomes. Table 2.6 presents the estimates of the variance-covariance matrix of initial latent variables. Since the standard deviation of all inputs is normalized to 1, covariances are the same as correlations. The estimates are derived from the most unrestricted model including the interaction terms of latent initial skill and investments, reported in model (4) - IV of Table 2.3.

Table 2.6: Initial Conditions

	$\ln \theta_0$	$\ln S_0$	$\ln T_0$	$\ln P_0$
$\ln \theta_0$	1.000			
	(0.000)			
$\ln S_0$	0.316	1.000		
	(0.0132)	(0.000)		
$\ln T_0$	0.0376	0.192	1.000	
	(0.0139)	(0.0136)	(0.000)	
$\ln P_0$	0.253	0.316	0.219	1.000
	(0.0134)	(0.0132)	(0.0136)	(0.000)

Source: CEPS
Notes: This table shows the correlation matrix of cognitive skills, school quality and parental investments. Standard errors are presented in parentheses. Results pertain to the model with complementarities and measurement errors corrected (see Model (4) - IV in Table 2.3).

Table 2.6 provides evidence of the assortative matching of students to schools: the correlation between initial skills and school quality is 0.32. Students with high initial cognitive skills have a higher chance of receiving better quality schooling. It also shows that parental investments in both private tutoring and parenting activities are positively correlated with their initial cognitive skills. The correlations are 0.04 and 0.25, respectively, indicating that parents are reinforcing initial skill advantages.

Turning to the relationship between school quality and parental investments, Table 2.6 indicates that parental investments and school quality are positively correlated. In particular, the correlation between tutoring investment and school quality is 0.19 and the correlation between parenting activities and school quality is 0.32. Students who are enrolled in a high quality school also tend to have higher parental investments, indicating that parents are reinforcing the high quality school inputs.

In sum, the positive correlations among students' initial skills, school quality, and parental investments accentuate the persistence of high-skill students.

Table 2.7: Inequality in School Quality and Parental Investments between Migrant and Local Students

	$\ln S_0$	$\ln P_0$	$\ln T_0$
Migrant Student	-0.0897***	-0.540***	-0.625***
	(0.0268)	(0.0356)	(0.0929)
No. of Students	5,176	5,176	5,176

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

Source: CEPS
Notes: Results pertain to the model with complementarities and measurement errors corrected (see Model (4) - IV in Table 2.3). The coefficients in columns (1)-(3) mean that the estimated average school quality $\ln S_0$, average parenting activities $\ln P_0$, average tutoring investment $\ln T_0$, for migrant students is 0.09 standard deviation, 0.54 standard deviation, 0.63 standard deviation lower than that for local students, respectively.

Table 2.8: Decomposition of Schools and Parents Inputs

(1) $E(\ln\theta \mid Migrant)$	(2) $E(\ln\theta \mid Local)$	(3) $E(\ln\theta \mid Migrant)$ $-E(\ln\theta \mid Local)$	(4) $E(\ln\theta \mid Migrant, S = E(S \mid Local))$ $-E(\ln\theta \mid Local)$	(5) $E(\ln\theta \mid Migrant, P = E(P \mid Local))$ $-E(\ln\theta \mid Local)$
0.41	0.50	-0.088	-0.051 (42%)	-0.073 (17%)

Source: CEPS

Notes: Column (4) shows the skill gap once the migrant students are matched with the average school quality of local students (and the percentage change with respect to column (3) in parenthesis). Column (5) shows the skill gap once the migrant students are matched with the average parental investments in parenting activities of local students (and the percentage change with respect to column (3) in parenthesis).

Inequality in Investments between Migrant and Local Students

I report the differences in investments from schools and parents between migrant and local students in Table 2.7. Column (1) shows how the estimated latent school quality differs between migrant and local students. The result is similar to the findings in the previous literature (Lu and Zhou, 2013; Chen and Feng, 2019) that compared to local students, migrant students attend schools with lower quality. Columns (2) and (3) indicate that migrant students also receive smaller amounts of parental investment in both private tutoring and parenting activities. However, the magnitude of the differences in parental investments is much larger than that of school quality.[15]

Decomposition Analysis

The importance of the various inputs of the model is quantified through a counterfactual decomposition analysis in this section.

Table 2.8 shows the effects for the main specification, reported in model (4) of Table 2.3. The first two columns present the average latent cognitive skills in the second period for migrant students and local students, respectively. The third column shows the skill gap between the two subgroups (the difference between columns (2) and (1)). The migrant-local skill gap is -0.088 standard deviation, with respect to the initial latent distributions (normalized to have standard deviation 1).

In columns (4) and (5), I counterfactually simulate changes in the initial conditions of the model. Column (4) shows the counterfactual skill gap predicted from the model if migrant students were provided the average school quality of the local students, and everything else remains the same. The percentage change of the skill gap relative to column (3) is reported in parenthesis. Column (5) shows the counterfactual skill gap predicted from the model if migrant students were provided with the average parental investment in parenting activities of the local students, and everything else remains the same. The results suggest that both school quality and parenting activities play an important role in

[15] Since all inputs are normalized to be mean 0 and variance 1, the coefficients are in standard deviations and can be compared across inputs.

explaining the migrant-local skill gap. School quality closes 42 percent of the skill gap, while parenting activities closes 17 percent of the skill gap.

2.6 Robustness Checks

2.6.1 Testing for Omitted Variable Bias

The main threat to the mean-independence assumption of production shock is the standard omitted variable bias: that the included latent variables for initial skills, parental and school investments are correlated with any omitted aspects of cognitive skill development.

Table 2.9: Test for Omitted Variable Bias

	(1) Residuals	(2) $\ln \theta_{i,1}$
Middle income family	-0.00330	0.358***
	(0.0285)	(0.0382)
Rich family	-0.0611	0.388***
	(0.0451)	(0.0606)
No. of Students	5,176	5,176

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

Source: CEPS
Notes: In column (1), the residuals of the specification with complementarities and measurement errors corrected (see Model (4) - IV in Table 2.3) are regressed on the category variable of household income level. The reference group is the poor families. In column (2), the new cognitive test scores ($\ln \theta_{i,1}$), are regressed on the same category variable.

In the first test of omitted variable bias, I use household income level as a proxy for the unobservable aspects as in Agostinelli et al. (2025). I regress the estimated residuals from the production function specification with complementarities (estimates reported in Model (4) - IV of Table 2.3) on family income level. Column (1) in Table 2.9 shows that household income level is unrelated with the residual variation of cognitive skill development. In contrast, column (2) in Table 2.9 shows that household income level is strongly positively correlated with students' cognitive skill in the second period. These findings are interpreted as a failure to reject the null hypothesis of no omitted variable bias.

Table 2.10: Model Fit of Average Skill by Student's Migration Status

$E(\ln \theta_{i,1})$	Data	Model Prediction	Diff.
Migrant	0.41	0.39	-0.02
	(0.023)	(0.017)	(0.018)
Local	0.50	0.51	0.01
	(0.016)	(0.010)	(0.012)

Standard errors in parentheses

Source: CEPS

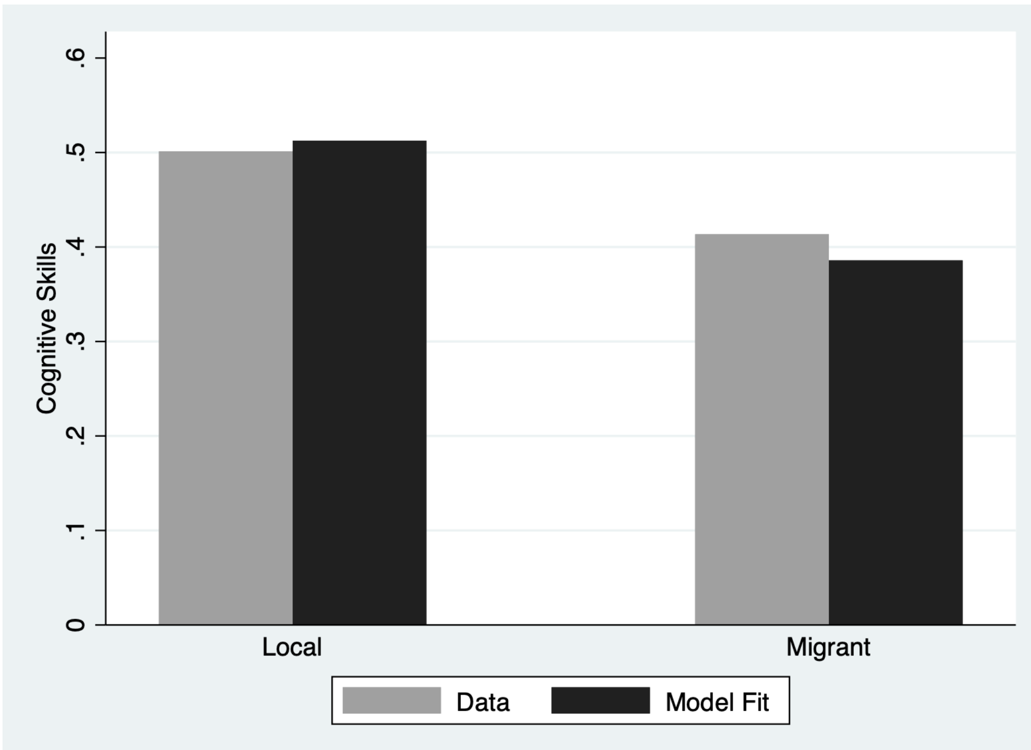
Notes: The numbers in the Data and Model Prediction columns represent the mean cognitive skills in the second period by the student's migration status (local versus migrant) in the data and predicted by the estimated model in Model (4) - IV of Table 2.3, respectively. The difference between the above two columns is shown in the last column. Migration status is not included in the estimated model; thus, model predictions for migrant and local students are outside of the model.

In the second test, I assess whether my model is sufficiently rich enough to capture the key aspects of cognitive skill development so that the model can make valid out-of-sample type predictions.[16] In Table 2.10 and Figure 2.3, I show that my production function specification based only on heterogeneity in children's initial skills, parental investments, and school quality is able to closely match the patterns of cognitive skill development

[16]The idea is essentially the same as testing whether the estimated model can predict "untargeted" moments in structural models.

by migration status.[17] This finding provides additional evidence that my model is not omitting important determinants of students' cognitive skill development.

Figure 2.3: Model Fit of Average Skill by Student's Migration Status



Source: CEPS
Notes: Each bar represents the mean cognitive skills in the second period by the student's migration status (local versus migrant) in the data (grey) and predicted by the estimated model in Model (4) - IV of Table 2.3 (black), respectively. Migration status is not included in the estimated model; thus, model predictions for migrant and local students are outside of the model.

2.6.2 Testing for Incidental Parameters Problem

In this section, I will address the potential issue of the incidental parameters problem when estimating equation (2.3) with the estimation algorithm in section 2.4.4. Recall that school input $\ln S_0$ is estimated as the average within-school residual in new cognitive skills. In other words, $\ln S_0$ can be seen as a school-specific fixed effect, which is the same for all

[17] This exercise would be uninformative if schools in the sample were completely segregated by migration status. If so, the model prediction would fit exactly the patterns by construction via the school effects. As shown in Figure B.3, schools in my sample are not fully segregated, but have some non-trivial variation in the fraction of migrant students.

students in the same school. Thus, the number of school fixed effects is the same as the number of schools. School fixed effects are the incidental parameters, of which the number increases with the number of schools, when the sample size increases as more schools are included in the sample. Recalling Step 1 of the estimation algorithm in section 2.4.4, only the observations of students who belong to a given school are informative in estimating the school fixed effect for that school. When the school size is small and fixed, school fixed effect itself cannot be consistently estimated. 10 percent of the schools in the baseline sample have fewer than 15 students, which causes concerns of inconsistent estimates for incidental parameters.

In addition, note that in the production function with complementarities in equation (2.3), the school input $\ln S_0$ enters the equation non-linearly, i.e., with an interaction term with initial cognitive skill, which means that school fixed effects cannot be eliminated by de-meaning or differencing within school. As a result, the estimation of productivity parameters (common parameters), which is a function of the school fixed effects, is also inconsistent, if the school fixed effects (incidental parameters) are inconsistently estimated.

The grouped fixed-effects estimator proposed in Bonhomme and Manresa (2015) assumes finite number of groups so that the number of fixed effects does not increase with sample size. By classifying the observations into finite groups, the number of observations used to estimate each fixed effect is larger and increases with sample size, so the estimates of fixed effects are consistent. Then, the productivity parameters in the production function can be consistently estimated.

As a robustness check for the estimated results in Model (4) - IV of Table 2.3, I employ the grouped fixed-effects estimator and extend the recursive estimation algorithm in section 2.4.4 in the following steps:

Step 0: Set the number of groups, N, i.e., $g \in \{1, ..., N\}$.

Step 1: Given the current guess of $\{\ln G_g\}_{g=1}^{N}$, the optimal group assignment of student i is

$$g(i) = \underset{g\in\{1,\ldots,N\}}{\text{argmin}} \sum_{j\in S(i)} (\tilde{Z}_{\theta,j,1} - \ln A - \gamma_1 \tilde{Z}_{\theta,j,0} - \gamma_2 \tilde{Z}_{T,j,m,0} - \gamma_3 \tilde{Z}_{P,j,m,0} - \gamma_4 \ln G_{g(i)}$$
$$\gamma_5 \tilde{Z}_{\theta,j,0} \tilde{Z}_{T,j,m,0} - \gamma_6 \tilde{Z}_{\theta,j,0} \tilde{Z}_{P,j,m,0} - \gamma_7 \tilde{Z}_{\theta,j,0} \ln G_{g(i)})^2$$

Step 2:

- **Step 2.1**: Given the current parameter guess, $\{\ln A^n, \{\gamma_k^n\}_{k=1}^7\}$, estimate the grouped fixed effects as the average within-group residual in new cognitive skills:

$$\ln G_{g(i)}^n = \frac{\begin{aligned}\sum_{j\in g(i)} \big(\tilde{Z}_{\theta,j,1} - \ln A^n - \gamma_1^n \tilde{Z}_{\theta,j,0} - \gamma_2^n \tilde{Z}_{T,j,m,0} - \gamma_3^n \tilde{Z}_{P,j,m,0} \\ - \gamma_5^n \tilde{Z}_{\theta,j,0} \tilde{Z}_{T,j,m,0} - \gamma_6^n \tilde{Z}_{\theta,j,0} \tilde{Z}_{P,j,m,0}\big)\end{aligned}}{\sum_{j\in S(i)} (\gamma_4^n + \gamma_7^n \tilde{Z}_{\theta,j,0})}$$

  where $g(i)$ is the set of students in the same group as student i.

- **Step 2.2**: Given the distribution of grouped fixed effects $\ln G_{g(i)}^n$ from Step 2.1, I estimate the cognitive skill production function in equation 2.3.

$$\begin{aligned}\tilde{Z}_{\theta,i,1} = \ln A^{n+1} &+ \gamma_1^{n+1} \tilde{Z}_{\theta,i,0} + \gamma_2^{n+1} \tilde{Z}_{T,i,m,0} + \gamma_3^{n+1} \tilde{Z}_{P,i,m,0} + \gamma_4^{n+1} \ln G_{g(i)}^n \\ &+ \gamma_5^{n+1} \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{T,i,m,0} + \gamma_6^{n+1} \tilde{Z}_{\theta,i,0} \times \tilde{Z}_{P,i,m,0} \\ &+ \gamma_7^{n+1} \tilde{Z}_{\theta,i,0} \times \ln G_{g(i)}^n + \kappa_{i,m,0}\end{aligned}$$

  where $\kappa_{i,m,0} = \eta_{i,0} - \gamma_2 \tilde{\epsilon}_{T,i,m,0} - \gamma_3 \tilde{\epsilon}_{P,i,m,0} - \gamma_5 \tilde{Z}_{\theta,i,m,0} \tilde{\epsilon}_{T,i,m,0} - \gamma_6 \tilde{Z}_{\theta,i,m,0} \tilde{\epsilon}_{P,i,m,0}$.

  Since $\kappa_{i,m,0}$ contains measurement errors, even with known grouped fixed effects, OLS estimation of the Step 2.2 equation would produce an inconsistent estimate of the structural parameters. Therefore, the equation in Step 2.2 is estimated via 2SLS estimation using the multiple excluded measures of parental investments ($Z_{T,i,m',0}$, $Z_{P,i,m',0}$) for some $m' \neq m$ as instrumental variables. This produces the $n + 1$ iteration of parameters, $\{\ln A^{n+1}, \{\gamma_k^{n+1}\}_{k=1}^7\}$, which can be used in Step 2.1. The inner iteration in Step 2 stops when the production function parameters converge.

The estimated grouped fixed effects $\ln G_{g(i)}$ will be used in Step 1 and the outer iteration of Steps 1 and 2 stops when the production function parameters and $\ln G_{g(i)}$ converge.

Table 2.11 reports the estimated results of equation (2.3) with grouped fixed-effects estimator. I experiment with different numbers of groups. Columns (1) to (4) show the results for 5 to 8 groups, respectively. The estimates are relatively stable. Compared to the estimates of the main specification in Model (4) - IV in Table 2.3, the difference is small, suggesting that the incidental parameter bias appears minimal.

Table 2.11: The Cognitive Skill Production Technology (Grouped Fixed Effects)

$\ln \theta_{i,1}$	(1)	(2)	(3)	(4)
$\ln \theta_{i,0}$	0.524***	0.477***	0.517***	0.514***
	(0.00464)	(0.0131)	(0.00983)	(0.00766)
$\ln G_{i,0}$	0.481***	0.538***	0.447***	0.419***
	(0.0106)	(0.0159)	(0.00693)	(0.00680)
$\ln T_{i,0}$	-0.0177	-0.0138	-0.0396***	-0.0392***
	(0.0104)	(0.00950)	(0.00441)	(0.00539)
$\ln P_{i,0}$	0.0943***	0.0785***	0.0816***	0.0821***
	(0.00556)	(0.00600)	(0.00944)	(0.00933)
$\ln \theta_{i,0} \times \ln G_{i,0}$	-0.179***	-0.225***	-0.162***	-0.147***
	(0.0120)	(0.0167)	(0.0117)	(0.00877)
$\ln \theta_{i,0} \times \ln T_{i,0}$	0.0000	0.0007	0.0106	0.0104
	(0.00907)	(0.00913)	(0.00901)	(0.00736)
$\ln \theta_{i,0} \times \ln P_{i,0}$	-0.0318**	-0.0283**	-0.0289	-0.0303*
	(0.0113)	(0.0101)	(0.0150)	(0.0150)
No. of Students	5,176	5,176	5,176	5,176
No. of Groups	5	6	7	8

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

Source: CEPS
Notes: All models control for students' gender, age, and number of siblings. All estimates are corrected for measurement errors.

2.7 Conclusion

This paper establishes a unified empirical framework that nests many key features of the child development literature and the education production function literature. The framework allows for both school and parental influences in cognitive skill development. In addition, unlike many previous papers assuming positive complementarity between existing skills and inputs of the production function, the signs of the complementarities between students' current cognitive skills and investments from schools and parents are unrestricted and estimated with the data using the model.

I find that school quality and parental investment in parenting activities are productive in building students' cognitive skills. However, the investment in private tutoring, although pervasive, is not productive in developing cognitive skills. In addition, I identify a negative complementarity between students' initial cognitive skills and both school quality and parental investment in parenting activities, suggesting that lower-skill students benefit more from an increase in school quality and parenting activities investment. Relating to the fact that migrant students generally have lower cognitive skills, policies that improve school quality and promote parental investment in parenting activities would be more beneficial to migrant students. The counterfactual policy experiments demonstrate that providing migrant students with either school quality or parenting activities investments of local students would substantially reduce the migrant-local skill gap.

Appendix A

Appendix for Chapter 1

A.1 Model Solutions

The model predicts that 1) the child's and parents' choices $\{L_{i,t}, M_{i,t}, S_{i,t}\}$ do not depend on state variable $\theta_{i,t}$; 2) conditional on parameters, parents' two choices, educational investment $M_{i,t}$ and discussion $S_{i,t}$, are independent; 3) child effort $L_{i,t}$ depends on parental discussion $S_{i,t}$ but not on parental educational investment $M_{i,t}$.

A.1.1 Child's Problem

Child i chooses $L_{i,t}$ after observing $S_{i,t}$ and $M_{i,t}$. Child's value function at period t:

$$V^c_{i,t}(\theta_{i,t}) = \max_{L_{i,t}|M_{i,t},S_{i,t}} \ln\theta_{i,t} + \lambda_{i,t}(1-\tau S_{i,t})\ln(\bar{L} - L_{i,t}) + \beta^c \mathbb{E}_{\eta_{i,t}}[V^c_{i,t+1}(\theta_{i,t+1})]$$

$$s.t. \quad \ln\theta_{i,t+1} = \ln A_i + \alpha_1 \ln\theta_{i,t} + \alpha_2 \ln L_{i,t} + \alpha_3 \ln M_{i,t} + \eta_{i,t}$$

where the expectation $\mathbb{E}_{\eta_{i,t}}$ is over production shock $\eta_{i,t}$ and $\mathbb{E}(\eta_{i,t}) = 0$.

Solutions. For all $t = 1, 2, .., T$, current and future optimal choices $\{L^*_{i,s}\}^T_{s=t}$ do not depend on $\theta_{i,t}$:

$$L^*_{i,t}(S_{i,t}) = \frac{\bar{L}\beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_{i,t}(1-\tau S_{i,t}) + \beta^c\Gamma^c_{t+1}\alpha_2}$$

where the sequence of $\{\Gamma_t^c\}_{t=1}^{T+1}$ is defined recursively and common across individuals:

$$\begin{aligned}
\Gamma_{T+1}^c &= \psi^c \\
\Gamma_T^c &= 1 + \beta^c \Gamma_{T+1}^c \alpha_1 \\
&... \\
\Gamma_t^c &= 1 + \beta^c \Gamma_{t+1}^c \alpha_1 \\
&... \\
\Gamma_1^c &= 1 + \beta^c \Gamma_2^c \alpha_1
\end{aligned}$$

Therefore, the value function $V_{i,t}^c(\theta_{i,t})$ is additively separable in state variable $\theta_{i,t}$ and current and future optimal choices $\{L_{i,s}^*\}_{s=t}^T$:

$$V_{i,t}^c(\theta_{i,t}) = V_{i,t}^{c1} + \sum_{s=t}^{T} (\beta^c)^{s-t} (V_{i,s}^{c2} + V_{i,s}^{c3}) \tag{A.1}$$

where

1. $V_{i,t}^{c1} = \Gamma_t^c \ln \theta_{i,t}$: current and future utility from $\theta_{i,t}$
2. $V_{i,s}^{c2} = \lambda_{i,t}(1 - \tau S_{i,s}) \ln(\bar{L} - L_{i,s}^*)$: flow utility from $L_{i,s}^*$ at period s, where $s = t, .., T$
3. $V_{i,s}^{c3} = \beta^c \Gamma_{s+1}^c (\ln A_i + \alpha_2 \ln L_{i,s}^* + \alpha_3 \ln M_{i,s})$: future utility from $L_{i,s}^*$ after period s, where $s = t, .., T$

Proof. At final period T, child i chooses $L_{i,T}$ after observing $S_{i,t}$ and $M_{i,T}$. Child's problem is:

$$V_{i,T}^c(\theta_{i,T}) = \max_{L_{i,T}|M_{i,T},S_{i,T}} \ln \theta_{i,T} + \lambda_{i,T}(1 - \tau S_{i,T}) \ln(\bar{L} - L_{i,T}) + \beta^c \mathbb{E}_{\eta_{i,T}}(\psi^c \ln \theta_{i,T+1})$$

$$s.t. \quad \ln \theta_{i,T+1} = \ln A_i + \alpha_1 \ln \theta_{i,T} + \alpha_2 \ln L_{i,T} + \alpha_3 \ln M_{i,T} + \eta_{i,T}$$

Substitute the production function into the value function:

$$V_{i,T}^{c}(\theta_{i,T}) = \max_{L_{i,T}|M_{i,T},S_{i,T}} \ln\theta_{i,T} + \lambda_{i,T}(1-\tau S_{i,T})\ln(\bar{L}-L_{i,T})$$
$$+ \beta^{c}\Gamma_{T+1}^{c}(\ln A_i + \alpha_1\ln\theta_{i,T} + \alpha_2\ln L_{i,T} + \alpha_3\ln M_{i,T})$$

Find the first-order condition for $L_{i,T}$:

$$\frac{\partial V_{i,T}^{c}}{\partial L_{i,T}} = -\frac{\lambda_{i,T}(1-\tau S_{i,T})}{\bar{L}-L_{i,T}} + \beta^{c}\Gamma_{T+1}^{c}\frac{\alpha_2}{L_{i,T}} = 0$$

Solve for optimal $L_{i,T}^{*}$:

$$L_{i,T}^{*}(S_{i,T}) = \frac{\bar{L}\beta^{c}\Gamma_{T+1}^{c}\alpha_2}{\lambda_{i,T}(1-\tau S_{i,T}) + \beta^{c}\Gamma_{T+1}^{c}\alpha_2}$$

Therefore, optimal choice $L_{i,T}^{*}$ does not depend on $\theta_{i,T}$ and the value function $V_{i,T}^{c}(\theta_{i,T})$ is additively separable in state variable $\theta_{i,T}$ and optimal choice $L_{i,T}^{*}$:

$$\begin{aligned}
V_{i,T}^{c}(\theta_{i,T}) =& \ln\theta_{i,T} + \lambda_{i,T}(1-\tau S_{i,T})\ln(\bar{L}-L_{i,T}^{*}) + \beta^{c}\mathbb{E}_{\eta_{i,T}}(\psi^{c}\ln\theta_{i,T+1})\\
=& \ln\theta_{i,T} + \lambda_{i,T}(1-\tau S_{i,T})\ln(\bar{L}-L_{i,T}^{*})\\
&+ \beta^{c}\Gamma_{T+1}^{c}(\ln A_i + \alpha_1\ln\theta_{i,T} + \alpha_2\ln L_{i,T}^{*} + \alpha_3\ln M_{i,T})\\
=&(1+\beta^{c}\Gamma_{T+1}^{c}\alpha_1)\ln\theta_{i,T} + \lambda_{i,T}(1-\tau S_{i,T})\ln(\bar{L}-L_{i,T}^{*})\\
&+ \beta^{c}\Gamma_{T+1}^{c}(\ln A_i + \alpha_2\ln L_{i,T}^{*} + \alpha_3\ln M_{i,T})\\
=&\Gamma_{T}^{c}\ln\theta_{i,T} + \lambda_{i,T}(1-\tau S_{i,T})\ln(\bar{L}-L_{i,T}^{*})\\
&+ \beta^{c}\Gamma_{T+1}^{c}(\ln A_i + \alpha_2\ln L_{i,T}^{*} + \alpha_3\ln M_{i,T})\\
=&V_{i,T}^{c1} + V_{i,T}^{c2} + V_{i,T}^{c3}
\end{aligned}$$

For $t = 1, 2, .., T-1$, prove by backward induction.

Suppose $\{L_{i,s}^{*}\}_{s=t+1}^{T}$ do not depend on $\theta_{i,t+1}$ and equation (A.1) holds in period $t+1$, it can be shown that $\{L_{i,s}^{*}\}_{s=t}^{T}$ do not depend on $\theta_{i,t}$ and equation (A.1) also holds in period t.

At period t, the child's problem:

$$V^c_{i,t}(\theta_{i,t}) = \max_{L_{i,t}|M_{i,t},S_{i,t}} \ln\theta_{i,t} + \lambda_{i,t}(1-\tau S_{i,t})\ln(\bar{L} - L_{i,t}) + \beta^c \mathbb{E}_{\eta_{i,t}}[V^c_{i,t+1}(\theta_{i,t+1})]$$
$$s.t. \quad \ln\theta_{i,t+1} = \ln A_i + \alpha_1 \ln\theta_{i,t} + \alpha_2 \ln L_{i,t} + \alpha_3 \ln M_{i,t} + \eta_{i,t}$$

Since future choices $\{L^*_{i,s}\}^T_{s=t+1}$ do not depend on $\theta_{i,t+1}$ or directly depend on current choice $L_{i,t}$, the choice of $L_{i,t}$ only impacts $V^{c1}_{i,t+1}$ in $V^c_{i,t+1}(\theta_{i,t+1})$ by changing $\theta_{i,t+1}$:

$$\begin{aligned} V^c_{i,t+1}(\theta_{i,t+1}|L_{i,t}) =& V^{c1}_{i,t+1}|_{L_{i,t}} + \sum_{s=t+1}^{T} (\beta^c)^{s-t-1}(V^{c2}_{i,s} + V^{c3}_{i,s}) \\ =& \Gamma^c_{t+1} \ln\theta_{i,t+1}|_{L_{i,t}} + \sum_{s=t+1}^{T} (\beta^c)^{s-t-1}(V^{c2}_{i,s} + V^{c3}_{i,s}) \end{aligned}$$

Find the first-order condition for $L_{i,t}$:

$$\begin{aligned} \frac{\partial V^c_{i,t}}{\partial L_{i,t}} = -\frac{\lambda_{i,t}(1-\tau S_{i,t})}{\bar{L} - L_{i,t}} + \beta^c \frac{\partial \mathbb{E}_{\eta_{i,t}}[V^c_{i,t+1}(\theta_{i,t+1}|L_{i,t})]}{\partial \ln\theta_{i,t+1}} \frac{\partial \ln\theta_{i,t+1}}{\partial L_{i,t}} = 0 \\ -\frac{\lambda_{i,t}(1-\tau S_{i,t})}{\bar{L} - L_{i,t}} + \beta^c \Gamma^c_{t+1} \frac{\alpha_2}{L_{i,t}} = 0 \end{aligned}$$

Solve for optimal $L^*_{i,t}$:

$$L^*_{i,t}(S_{i,t}) = \frac{\bar{L}\beta^c\Gamma^c_{t+1}\alpha_2}{\lambda_{i,t}(1-\tau S_{i,t}) + \beta^c\Gamma^c_{t+1}\alpha_2}$$

Therefore, optimal choice $L^*_{i,t}$ does not depend on $\theta_{i,t}$. Since $\{L^*_{i,s}\}^T_{s=t+1}$ do not depend on $\theta_{i,t+1}$, $\{L^*_{i,s}\}^T_{s=t+1}$ do not depend on $\theta_{i,t}$. In sum, $\{L^*_{i,s}\}^T_{s=t}$ do not depend on $\theta_{i,t}$ and equation (A.1) is derived as follows:

$$
\begin{aligned}
V^c_{i,t}(\theta_{i,t}) =& \ln \theta_{i,t} + \lambda_{i,t}(1 - \tau S_{i,t}) \ln(\bar{L} - L^*_{i,t}) + \beta^c \mathbb{E}_{\eta_{i,t}}[V^c_{i,t+1}(\theta_{i,t+1}|L^*_{i,t})] \\
=& V^{c2}_{i,t} + \ln \theta_{i,t} + \beta^c(\mathbb{E}_{\eta_{i,t}}[V^{c1}_{i,t+1}|L^*_{i,t}] + \sum_{s=t+1}^{T} (\beta^c)^{s-t-1}(V^{c2}_{i,s} + V^{c3}_{i,s})) \\
=& V^{c1}_{i,t} + V^{c2}_{i,t} + V^{c3}_{i,t} + \sum_{s=t+1}^{T} (\beta^c)^{s-t}(V^{c2}_{i,s} + V^{c3}_{i,s}) \\
=& V^{c1}_{i,t} + \sum_{s=t}^{T} (\beta^c)^{s-t}(V^{c2}_{i,s} + V^{c3}_{i,s})
\end{aligned}
$$

as

$$
\begin{aligned}
V^{c1}_{i,t} + V^{c3}_{i,t} =& \Gamma^c_t \ln \theta_{i,t} + \beta^c \Gamma^c_{t+1}(\ln A_i + \alpha_2 \ln L^*_{i,t} + \alpha_3 \ln M_{i,t}) \\
=& (1 + \beta^c \Gamma^c_{t+1} \alpha_1) \ln \theta_{i,t} + \beta^c \Gamma^c_{t+1}(\ln A_i + \alpha_2 \ln L^*_{i,t} + \alpha_3 \ln M_{i,t}) \\
=& \ln \theta_{i,t} + \beta^c \Gamma^c_{t+1}(\alpha_1 \ln \theta_{i,t} + \ln A_i + \alpha_2 \ln L^*_{i,t} + \alpha_3 \ln M_{i,t}) \\
=& \ln \theta_{i,t} + \beta^c \Gamma^c_{t+1} \mathbb{E}_{\eta_{i,t}}[\ln \theta_{i,t+1}|L^*_{i,t}] \\
=& \ln \theta_{i,t} + \beta^c \mathbb{E}_{\eta_{i,t}}[V^{c1}_{i,t+1}|L^*_{i,t}]
\end{aligned}
$$

$\square$

Given the solutions to the child's problem, the change in the child's utility due to parental discussion is calculated below.

$$
\begin{aligned}
& V_{i,t}^{c}(\theta_{i,t}|S_{i,t}=1) - V_{i,t}^{c}(\theta_{i,t}|S_{i,t}=0) \\
=& [\lambda_{i,t}(1-\tau)\ln(\bar{L}-L_{i,t}^{*}(S_{i,t}=1)) - \lambda_{i,t}\ln(\bar{L}-L_{i,t}^{*}(S_{i,t}=0))] \\
& + [\beta^{c}\Gamma_{t+1}^{c}\alpha_2 \ln L_{i,t}^{*}(S_{i,t}=1) - \beta^{c}\Gamma_{t+1}^{c}\alpha_2 \ln L_{i,t}^{*}(S_{i,t}=0)] \\
=& \lambda_{i,t}\ln\frac{\bar{L}-L_{i,t}^{*}(S_{i,t}=1)}{\bar{L}-L_{i,t}^{*}(S_{i,t}=0)} - \lambda_{i,t}\tau\ln(\bar{L}-L_{i,t}^{*}(S_{i,t}=1)) + \beta^{c}\Gamma_{t+1}^{c}\alpha_2\ln\frac{L_{i,t}^{*}(S_{i,t}=1)}{L_{i,t}^{*}(S_{i,t}=0)} \\
=& \underbrace{\lambda_{i,t}\ln(\bar{L} - \frac{\tau\beta^{c}\Gamma_{t+1}^{c}\alpha_2}{\lambda_{i,t}(1-\tau)+\beta^{c}\Gamma_{t+1}^{c}\alpha_2})}_{\text{cost from greater effort}} - \underbrace{\lambda_{i,t}\tau\ln\frac{\lambda_{i,t}(1-\tau)}{\lambda_{i,t}(1-\tau)+\beta^{c}\Gamma_{t+1}^{c}\alpha_2}}_{\text{cost from decreased value of leisure}} \\
& + \underbrace{\beta^{c}\Gamma_{t+1}^{c}\alpha_2\ln\frac{\lambda_{i,t}+\beta^{c}\Gamma_{t+1}^{c}\alpha_2}{\lambda_{i,t}(1-\tau)+\beta^{c}\Gamma_{t+1}^{c}\alpha_2}}_{\text{benefit from higher skill}}
\end{aligned}
$$

as

$$\bar{L} - L_{i,t}^{*}(S_{i,t}) = \frac{\bar{L}\lambda_{i,t}(1-\tau S_{i,t})}{\lambda_{i,t}(1-\tau S_{i,t})+\beta^{c}\Gamma_{t+1}^{c}\alpha_2}$$

and

$$
\begin{aligned}
\frac{\bar{L}-L_{i,t}^{*}(S_{i,t}=1)}{\bar{L}-L_{i,t}^{*}(S_{i,t}=0)} &= \frac{(1-\tau)(\lambda_{i,t}+\beta^{c}\Gamma_{t+1}^{c}\alpha_2)}{\lambda_{i,t}(1-\tau)+\beta^{c}\Gamma_{t+1}^{c}\alpha_2} \\
&= 1 - \frac{\tau\beta^{c}\Gamma_{t+1}^{c}\alpha_2}{\lambda_{i,t}(1-\tau)+\beta^{c}\Gamma_{t+1}^{c}\alpha_2}
\end{aligned}
$$

A.1.2 Parents' Problem

Given state variable $\theta_{i,t}$, income $Y_{i,t}$, and preference shocks to consumption and discussion cost, parents i make decisions on $S_{i,t} \in \{0,1\}$ and $M_{i,t}$ conditional on the child's response $L_{i,t}^{*}(S_{i,t})$. Parents' value function at period t:

$$
\begin{aligned}
V_{i,t}^{p}(\theta_{i,t}) = \max_{M_{i,t},S_{i,t}|L_{i,t}^{*}(S_{i,t})} & \ln\theta_{i,t} + \delta_{i,t}\ln(Y_{i,t}-M_{i,t}) - \xi_{i,t}S_{i,t} + \beta^{p}\mathbb{E}_{\eta_{i,t}}[V_{i,t+1}^{p}(\theta_{i,t+1})] \\
s.t. \quad & \ln\theta_{i,t+1} = \ln A_i + \alpha_1\ln\theta_{i,t} + \alpha_2\ln L_{i,t} + \alpha_3\ln M_{i,t} + \eta_{i,t}
\end{aligned}
$$

where the expectation $\mathbb{E}_{\eta_{i,t}}$ is over production shock $\eta_{i,t}$ and $\mathbb{E}(\eta_{i,t}) = 0$.

Solutions. For all $t = 1, 2, .., T$, current and future optimal choices $\{M^*_{i,s}, S^*_{i,s}\}_{s=t}^T$ do not depend on $\theta_{i,t}$:

1. Optimal parental monetary investment $M^*_{i,t}$:

$$M^*_{i,t} = \frac{\beta^p \Gamma^p_{t+1} \alpha_3}{\delta_{i,t} + \beta^p \Gamma^p_{t+1} \alpha_3} Y_{i,t}$$

   where the sequence of $\{\Gamma^p_t\}_{t=1}^{T+1}$ is defined recursively and common across individuals:

$$\begin{aligned}
\Gamma^p_{T+1} &= \psi^p \\
\Gamma^p_T &= 1 + \beta^p \Gamma^p_{T+1} \alpha_1 \\
&\dots \\
\Gamma^p_t &= 1 + \beta^p \Gamma^p_{t+1} \alpha_1 \\
&\dots \\
\Gamma^p_1 &= 1 + \beta^p \Gamma^p_2 \alpha_1
\end{aligned}$$

2. Optimal discussion decision $S^*_{i,t} = \mathbf{1}\{V^p_{i,t}(\theta_{i,t}|S_{i,t} = 1) > V^p_{i,t}(\theta_{i,t}|S_{i,t} = 0)\}$, where $\mathbf{1}\{\cdot\}$ is an indicator function and the difference in conditional values is calculated below.

$$V^p_{i,t}(\theta_{i,t}|S_{i,t} = 1) - V^p_{i,t}(\theta_{i,t}|S_{i,t} = 0) = -\xi_{i,t} + \beta^p \Gamma^p_{t+1} \alpha_2 \ln \frac{\lambda_i + \beta^c \Gamma^c_{t+1} \alpha_2}{\lambda_i(1-\tau) + \beta^c \Gamma^c_{t+1} \alpha_2}$$

   Define $B_{i,t}$ as the benefit of discussion.

$$B_{i,t} = \beta^p \Gamma^p_{t+1} \alpha_2 \ln \frac{\lambda_i + \beta^c \Gamma^c_{t+1} \alpha_2}{\lambda_i(1-\tau) + \beta^c \Gamma^c_{t+1} \alpha_2}$$

   Let $q_{i,t}$ be the probability of parents i choosing to discuss performance with the child in period t when the benefit outweighs the cost: $q_{i,t} = \mathbb{P}(S^*_{i,t} = 1) = \mathbb{P}(\xi_{i,t} < B_{i,t})$.

Given the distribution of discussion cost, the probability $q_{i,t}$ is calculated below.

$$q_{i,t} = \Phi(\frac{\ln B_{i,t} - \ln \xi_i}{\sigma_\xi})$$

Therefore, the value function $V^p_{i,t}(\theta_{i,t})$ is additively separable in state variable $\theta_{i,t}$ and current and future optimal choices $\{M^*_{i,s}, S^*_{i,s}\}^T_{s=t}$:

$$V^p_{i,t}(\theta_{i,t}) = V^{p1}_{i,t} + \sum_{s=t}^{T}(\beta^p)^{s-t}(V^{p2}_{i,s} + V^{p3}_{i,s}) \tag{A.2}$$

where

1. $V^{p1}_{i,t} = \Gamma^p_t \ln \theta_{i,t}$: current and future utility from $\theta_{i,t}$
2. $V^{p2}_{i,s} = \delta_{i,t}\ln(Y_{i,s} - M^*_{i,s}) - \xi_{i,s}S^*_{i,s}$: flow utility from $\{M^*_{i,s}, S^*_{i,s}\}$ in period s, where $s = t, .., T$
3. $V^{p3}_{i,s} = \beta^p\Gamma^p_{s+1}(\ln A_i + \alpha_2 \ln L^*_{i,s}(S^*_{i,s}) + \alpha_3 \ln M^*_{i,s})$: future utility from $\{M^*_{i,s}, S^*_{i,s}\}$ after period s, where $s = t, .., T$

Proof. At final period T, conditional on the child's response $L^*_{i,T}(S_{i,T})$, parents' problem is:

$$V^p_{i,T}(\theta_{i,T}) = \max_{M_{i,T}, S_{i,T}|L^*_{i,T}(S_{i,T})} \ln \theta_{i,T} + \delta_{i,T}\ln(Y_{i,T} - M_{i,T}) - \xi_{i,T}S_{i,T} + \beta^p\mathbb{E}_{\eta_{i,T}}[\psi^p \ln \theta_{i,T+1}]$$

$$s.t. \quad \ln \theta_{i,T+1} = \ln A_i + \alpha_1 \ln \theta_{i,T} + \alpha_2 \ln L^*_{i,T}(S_{i,T}) + \alpha_3 \ln M_{i,T} + \eta_{i,T}$$

Substitute the production function into the value function:

$$\begin{aligned} V^p_{i,T}(\theta_{i,T}) = \max_{M_{i,T}, S_{i,T}|L^*_{i,T}(S_{i,T})} & \ln \theta_{i,T} + \delta_{i,T}\ln(Y_{i,T} - M_{i,T}) - \xi_{i,T}S_{i,T} \\ & + \beta^p\Gamma^p_{T+1}(\ln A_i + \alpha_1 \ln \theta_{i,T} + \alpha_2 \ln L^*_{i,T}(S_{i,T}) + \alpha_3 \ln M_{i,T}) \end{aligned}$$

F.O.C. of $M_{i,T}$:

$$\frac{\delta_{i,T}}{Y_{i,T} - M_{i,T}} = \beta^p \Gamma^p_{T+1} \frac{\alpha_3}{M_{i,T}}$$

Solve for optimal $M^*_{i,T}$:

$$M^*_{i,T} = \frac{\beta^p \Gamma^p_{T+1} \alpha_3}{\delta_{i,T} + \beta^p \Gamma^p_{T+1} \alpha_3} Y_{i,T}$$

Given optimal $M^*_{i,T}$, the conditional values $V^p_{i,T}(\theta_{i,T}|S_{i,T} = j)$, $j \in \{0, 1\}$:

$$\begin{aligned} V^p_{i,T}(\theta_{i,T}|S_{i,T} = j) = {} & \ln \theta_{i,T} + \delta_{i,T} \ln(Y_{i,T} - M^*_{i,T}) - \xi_{i,T} S_{i,T} \\ & + \beta^p \Gamma^p_{T+1} \mathbb{E}_{\eta_{i,T}} [\ln \theta_{i,T+1} | S_{i,T} = j] \end{aligned}$$

where $\mathbb{E}_{\eta_{i,T}}[\ln \theta_{i,T+1}|S_{i,T} = j] = \ln A_i + \alpha_1 \ln \theta_{i,T} + \alpha_2 \ln L^*_{i,T}(S_{i,T} = j) + \alpha_3 \ln M^*_{i,T}$ and

$$L^*_{i,T}(S_{i,T} = j) = \frac{\beta^c \Gamma^c_{T+1} \alpha_2}{\lambda_{i,T}(1 - \tau S_{i,T}) + \beta^c \Gamma^c_{T+1} \alpha_2}$$

Solve for optimal choice $S^*_{i,T} = \mathbf{1}\{V^p_{i,T}(\theta_{i,T}|S_{i,T} = 1) > V^p_{i,T}(\theta_{i,T}|S_{i,T} = 0)\}$, where $\mathbf{1}\{\cdot\}$ is an indicator function and the difference in conditional values is calculated below.

$$\begin{aligned} & V^p_{i,T}(\theta_{i,T}|S_{i,T} = 1) - V^p_{i,T}(\theta_{i,T}|S_{i,T} = 0) \\ = & - \xi_{i,T} + \beta^p \Gamma^p_{T+1} (\mathbb{E}_{\eta_{i,T}} [\ln \theta_{i,T+1}|S_{i,T} = 1] - \mathbb{E}_{\eta_{i,T}} [\ln \theta_{i,T+1}|S_{i,T} = 0]) \\ = & - \xi_{i,T} + \beta^p \Gamma^p_{T+1} \alpha_2 (\ln L^*_{i,T}(S_{i,T} = 1) - \ln L^*_{i,T}(S_{i,T} = 0)) \\ = & - \xi_{i,T} + \beta^p \Gamma^p_{T+1} \alpha_2 \ln \frac{\lambda_{i,T} + \beta^c \Gamma^c_{T+1} \alpha_2}{\lambda_{i,T}(1 - \tau) + \beta^c \Gamma^c_{T+1} \alpha_2} \end{aligned}$$

Therefore, optimal choices $\{M^*_{i,T}, S^*_{i,T}\}$ do not depend on $\theta_{i,T}$ and the value function

$V^p_{i,T}(\theta_{i,T})$ is additively separable in state variable $\theta_{i,T}$ and optimal choices $\{M^*_{i,T}, S^*_{i,T}\}$:

$$\begin{aligned}
V^p_{i,T}(\theta_{i,T}) =& \ln\theta_{i,T} + \delta_{i,T}\ln(Y_{i,T} - M^*_{i,T}) - \xi_{i,T}S^*_{i,T} + \beta^p\mathbb{E}_{\eta_{i,T}}[\psi^p \ln\theta_{i,T+1}] \\
=& \ln\theta_{i,T} + \delta_{i,T}\ln(Y_{i,T} - M^*_{i,T}) - \xi_{i,T}S^*_{i,T} \\
& + \beta^p\Gamma^p_{T+1}(\ln A_i + \alpha_1\ln\theta_{i,T} + \alpha_2\ln L^*_{i,T}(S^*_{i,T}) + \alpha_3\ln M^*_{i,T}) \\
=&(1 + \beta^p\Gamma^p_{T+1}\alpha_1)\ln\theta_{i,T} + \delta_{i,T}\ln(Y_{i,T} - M^*_{i,T}) - \xi_{i,T}S^*_{i,T} \\
& + \beta^p\Gamma^p_{T+1}(\ln A_i + \alpha_2\ln L^*_{i,T}(S^*_{i,T}) + \alpha_3\ln M^*_{i,T}) \\
=&\Gamma^p_T\ln\theta_{i,T} + \delta_{i,T}\ln(Y_{i,T} - M^*_{i,T}) - \xi_{i,T}S^*_{i,T} \\
& + \beta^p\Gamma^p_{T+1}(\ln A_i + \alpha_2\ln L^*_{i,T}(S^*_{i,T}) + \alpha_3\ln M^*_{i,T}) \\
=&V^{p1}_{i,T} + V^{p2}_{i,T} + V^{p3}_{i,T}
\end{aligned}$$

For $t = 1, 2, .., T-1$, prove by backward induction.

Suppose $\{M^*_{i,s}, S^*_{i,s}\}^T_{s=t+1}$ do not depend on $\theta_{i,t+1}$ and equation (A.2) holds in period $t+1$, it can be shown that $\{M^*_{i,s}, S^*_{i,s}\}^T_{s=t}$ do not depend on $\theta_{i,t}$ and equation (A.2) also holds in period t.

At period t, conditional on the child's response $L^*_{i,t}(S_{i,t})$, parents' problem:

$$\begin{aligned}
V^p_{i,t}(\theta_{i,t}) = \max_{M_{i,t}, S_{i,t}|L^*_{i,t}(S_{i,t})} & \ln\theta_{i,t} + \delta_{i,t}\ln(Y_{i,t} - M_{i,t}) - \xi_{i,t}S_{i,t} \\
& + \beta^p\mathbb{E}_{\eta_{i,t}}[V^p_{i,t+1}(\theta_{i,t+1}|M_{i,t}, S_{i,t})] \\
s.t. \quad & \ln\theta_{i,t+1} = \ln A_i + \alpha_1\ln\theta_{i,t} + \alpha_2\ln L^*_{i,t}(S_{i,t}) + \alpha_3\ln M_{i,t} + \eta_{i,t}
\end{aligned}$$

Since future choices $\{M^*_{i,s}, S^*_{i,s}\}^T_{s=t+1}$ do not depend on $\theta_{i,t+1}$ or directly depend on current choice $\{M_{i,t}, S_{i,t}\}$ [1], the choice of $\{M_{i,t}, S_{i,t}\}$ only impacts $V^{p1}_{i,t+1}$ in $V^p_{i,t+1}(\theta_{i,t+1})$ by

[1]Examples of direct dependence of future choices on current choices: 1) if households can save and borrow, $\{M^*_{i,s}\}^T_{s=t+1}$ is directly affected by $M_{i,t}$; 2) if discussion cost $\xi_{i,t}$ is serially correlated, $\{S^*_{i,s}\}^T_{s=t+1}$ is correlated with $S_{i,t}$.

changing $\theta_{i,t+1}$:

$$\begin{aligned} V^p_{i,t+1}(\theta_{i,t+1}|M_{i,t},S_{i,t}) =& V^{p1}_{i,t+1}|_{M_{i,t},S_{i,t}} + \sum_{s=t+1}^{T}(\beta^p)^{s-t-1}(V^{p2}_{i,s}+V^{p3}_{i,s}) \\ =& \Gamma^p_{t+1}\ln\theta_{i,t+1}|_{M_{i,t},S_{i,t}} + \sum_{s=t+1}^{T}(\beta^p)^{s-t-1}(V^{p2}_{i,s}+V^{p3}_{i,s}) \end{aligned}$$

F.O.C. for $M_{i,t}$:

$$\begin{aligned} \frac{\partial V^p_{i,t}}{\partial M_{i,t}} = -\frac{\delta_{i,t}}{Y_{i,t}-M_{i,t}} + \beta^p \frac{\partial \mathbb{E}_{\eta_{i,t}}[V^p_{i,t+1}(\theta_{i,t+1}|M_{i,t},S_{i,t})]}{\partial \ln\theta_{i,t+1}}\frac{\partial \ln\theta_{i,t+1}}{\partial M_{i,t}} &= 0 \\ -\frac{\delta_{i,t}}{Y_{i,t}-M_{i,t}} + \beta^p\Gamma^p_{t+1}\frac{\alpha_3}{M_{i,t}} &= 0 \end{aligned}$$

Solve for optimal $M^*_{i,t}$:

$$M^*_{i,t} = \frac{\beta^p\Gamma^p_{t+1}\alpha_3}{\delta_{i,t}+\beta^p\Gamma^p_{t+1}\alpha_3}Y_{i,t}$$

Given the optimal choice of $M^*_{i,t}$, the conditional value $V^p_{i,t}(\theta_{i,t}|S_{i,t}=j)$, $j\in\{0,1\}$:

$$V^p_{i,t}(\theta_{i,t}|S_{i,t}=j) = \ln\theta_{i,t} + \delta_{i,t}\ln(Y_{i,t}-M^*_{i,t}) - \xi_{i,t}S_{i,t} + \beta^p\mathbb{E}_{\eta_{i,t}}[V^p_{i,t+1}(\theta_{i,t+1}|M^*_{i,t},S_{i,t}=j)]$$

where

$$\begin{aligned} \mathbb{E}_{\eta_{i,t}}[V^p_{i,t+1}(\theta_{i,t+1}|M^*_{i,t},S_{i,t}=j)] =& \mathbb{E}_{\eta_{i,t}}[V^{p1}_{i,t+1}|M^*_{i,t},S_{i,t}=j] + \sum_{s=t+1}^{T}(\beta^p)^{s-t-1}(V^{p2}_{i,s}+V^{p3}_{i,s}) \\ =& \Gamma^p_{t+1}\mathbb{E}_{\eta_{i,t}}[\ln\theta_{i,t+1}|M^*_{i,t},S_{i,t}=j] \\ &+ \sum_{s=t+1}^{T}(\beta^p)^{s-t-1}(V^{p2}_{i,s}+V^{p3}_{i,s}) \end{aligned}$$

and

$$\mathbb{E}_{\eta_{i,t}}[\ln\theta_{i,t+1}|M^*_{i,t},S_{i,t}=j] = \ln A_i + \alpha_1\ln\theta_{i,t} + \alpha_2\ln L^*_{i,t}(S_{i,t}=j) + \alpha_3\ln M^*_{i,t}$$

and

$$L_{i,t}^*(S_{i,t}) = \frac{\beta^c \Gamma_{t+1}^c \alpha_2}{\lambda_{i,t}(1 - \tau S_{i,t}) + \beta^c \Gamma_{t+1}^c \alpha_2}$$

Solve for optimal choice $S_{i,t}^* = \mathbf{1}\{V_{i,t}^p(\theta_{i,t}|S_{i,t} = 1) > V_{i,t}^p(\theta_{i,t}|S_{i,t} = 0)\}$, where $\mathbf{1}\{\cdot\}$ is an indicator function and the difference in conditional values is calculated below.

$$\begin{aligned}
&V_{i,t}^p(\theta_{i,t}|S_{i,t} = 1) - V_{i,t}^p(\theta_{i,t}|S_{i,t} = 0) \\
=& - \xi_{i,t} + \beta^p(\mathbb{E}_{\eta_{i,t}}[V_{i,t+1}^p(\theta_{i,t+1}|M_{i,t}^*, S_{i,t} = 1)] - \mathbb{E}_{\eta_{i,t}}[V_{i,t+1}^p(\theta_{i,t+1}|M_{i,t}^*, S_{i,t} = 0)]) \\
=& - \xi_{i,t} + \beta^p \Gamma_{t+1}^p(\mathbb{E}_{\eta_{i,t}}[\ln \theta_{i,t+1}|M_{i,t}^*, S_{i,t} = 1] - \mathbb{E}_{\eta_{i,t}}[\ln \theta_{i,t+1}|M_{i,t}^*, S_{i,t} = 0]) \\
=& - \xi_{i,t} + \beta^p \Gamma_{t+1}^p \alpha_2(\ln L_{i,t}^*(S_{i,t} = 1) - \ln L_{i,t}^*(S_{i,t} = 0)) \\
=& - \xi_{i,t} + \beta^p \Gamma_{t+1}^p \alpha_2 \ln \frac{\lambda_{i,t} + \beta^c \Gamma_{t+1}^c \alpha_2}{\lambda_{i,t}(1 - \tau) + \beta^c \Gamma_{t+1}^c \alpha_2}
\end{aligned}$$

Define $B_{i,t}$ as the benefit of discussion.

$$B_{i,t} = \beta^p \Gamma_{t+1}^p \alpha_2 \ln \frac{\lambda_{i,t} + \beta^c \Gamma_{t+1}^c \alpha_2}{\lambda_{i,t}(1 - \tau) + \beta^c \Gamma_{t+1}^c \alpha_2}$$

Let $q_{i,t}$ be the probability of parents i choosing to discuss performance with the child in period t when the benefit outweighs the cost: $q_{i,t} = \mathbb{P}(S_{i,t}^* = 1) = \mathbb{P}(\xi_{i,t} < B_{i,t})$. Given the distribution of discussion cost, the probability $q_{i,t}$ is calculated below.

$$q_{i,t} = \Phi(\frac{\ln B_{i,t} - \ln \xi_i}{\sigma_\xi})$$

Therefore, optimal choices $\{M_{i,t}^*, S_{i,t}^*\}$ do not depend on $\theta_{i,t}$. Since $\{M_{i,s}^*, S_{i,s}^*\}_{s=t+1}^T$ do not depend on $\theta_{i,t+1}$, $\{M_{i,s}^*, S_{i,s}^*\}_{s=t+1}^T$ do not depend on $\theta_{i,t}$. In sum, $\{M_{i,s}^*, S_{i,s}^*\}_{s=t}^T$ do not depend on $\theta_{i,t}$ and equation (A.2) is derived as follows:

$$\begin{aligned}
V_{i,t}^{p}(\theta_{i,t}) =& \ln\theta_{i,t} + \delta_{i,t}\ln(Y_{i,t} - M_{i,t}^{*}) - \xi_{i,t}S_{i,t}^{*} + \beta^{p}\mathbb{E}_{\eta_{i,t}}[V_{i,t+1}^{p}(\theta_{i,t+1}|M_{i,t}^{*}, S_{i,t}^{*})] \\
=& V_{i,t}^{p2} + \ln\theta_{i,t} + \beta^{p}(\mathbb{E}_{\eta_{i,t}}[V_{i,t+1}^{p1}|M_{i,t}^{*}, S_{i,t}^{*}] + \sum_{s=t+1}^{T}(\beta^{p})^{s-t-1}(V_{i,s}^{p2} + V_{i,s}^{p3})) \\
=& V_{i,t}^{p1} + V_{i,t}^{p2} + V_{i,t}^{p3} + \sum_{s=t+1}^{T}(\beta^{p})^{s-t}(V_{i,s}^{p2} + V_{i,s}^{p3}) \\
=& V_{i,t}^{p1} + \sum_{s=t}^{T}(\beta^{p})^{s-t}(V_{i,s}^{p2} + V_{i,s}^{p3})
\end{aligned}$$

as

$$\begin{aligned}
V_{i,t}^{p1} + V_{i,t}^{p3} =& \Gamma_{t}^{p}\ln\theta_{i,t} + \beta^{p}\Gamma_{t+1}^{p}(\ln A_{i} + \alpha_{2}\ln L_{i,t}^{*}(S_{i,t}^{*}) + \alpha_{3}\ln M_{i,t}^{*}) \\
=& (1 + \beta^{p}\Gamma_{t+1}^{p}\alpha_{1})\ln\theta_{i,t} + \beta^{p}\Gamma_{t+1}^{p}(\ln A_{i} + \alpha_{2}\ln L_{i,t}^{*}(S_{i,t}^{*}) + \alpha_{3}\ln M_{i,t}^{*}) \\
=& \ln\theta_{i,t} + \beta^{p}\Gamma_{t+1}^{p}(\alpha_{1}\ln\theta_{i,t} + \ln A_{i} + \alpha_{2}\ln L_{i,t}^{*}(S_{i,t}^{*}) + \alpha_{3}\ln M_{i,t}^{*}) \\
=& \ln\theta_{i,t} + \beta^{p}\Gamma_{t+1}^{p}\mathbb{E}_{\eta_{i,t}}[\ln\theta_{i,t+1}|M_{i,t}^{*}, S_{i,t}^{*}] \\
=& \ln\theta_{i,t} + \beta^{p}\mathbb{E}_{\eta_{i,t}}[V_{i,t+1}^{p1}|M_{i,t}^{*}, S_{i,t}^{*}]
\end{aligned}$$

□

A.2 Data and Estimation

A.2.1 Identification and Estimation

Arellano-Bond Moment Conditions

The moment conditions described in equation (1.11) is valid as shown below.

$$\begin{aligned}
\mathbb{E}[\tilde{z}^{j'}_{i,t-s}\nu^{j}_{i,t}] = \mathbb{E}[(\ln\theta_{i,t-s} + \tilde{\varepsilon}^{j'}_{i,t-s})(\eta_{i,t} + \tilde{\varepsilon}^{j}_{i,t+1} - \alpha_1\tilde{\varepsilon}^{j}_{i,t})] = 0 \\
\text{as} \quad & \mathbb{E}(\ln\theta_{i,t-s}\eta_{i,t}) = 0 \\
& \mathbb{E}(\ln\theta_{i,t-s}(\tilde{\varepsilon}^{j}_{i,t+1} - \alpha_1\tilde{\varepsilon}^{j}_{i,t})) = 0 \\
& \mathbb{E}(\tilde{\varepsilon}^{j'}_{i,t-s}\eta_{i,t}) = 0 \\
& \mathbb{E}(\tilde{\varepsilon}^{j'}_{i,t-s}(\tilde{\varepsilon}^{j}_{i,t+1} - \alpha_1\tilde{\varepsilon}^{j}_{i,t})) = 0 \quad \text{for } s \geq 1
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\tilde{z}^{j'}_{i,t-s}\nu^{j}_{i,t-1}] = \mathbb{E}[(\ln\theta_{i,t-s} + \tilde{\varepsilon}^{j'}_{i,t-s})(\eta_{i,t-1} + \tilde{\varepsilon}^{j}_{i,t} - \alpha_1\tilde{\varepsilon}^{j}_{i,t-1})] = 0 \\
\text{as} \quad & \mathbb{E}(\ln\theta_{i,t-s}\eta_{i,t-1}) = 0 \\
& \mathbb{E}(\ln\theta_{i,t-s}(\tilde{\varepsilon}^{j}_{i,t} - \alpha_1\tilde{\varepsilon}^{j}_{i,t-1})) = 0 \\
& \mathbb{E}(\tilde{\varepsilon}^{j'}_{i,t-s}\eta_{i,t-1}) = 0 \\
& \mathbb{E}(\tilde{\varepsilon}^{j'}_{i,t-s}(\tilde{\varepsilon}^{j}_{i,t} - \alpha_1\tilde{\varepsilon}^{j}_{i,t-1})) = 0 \quad \text{for } s \geq 1
\end{aligned}$$

Estimation Algorithm for Preference Parameters

Given $\{\beta^p, \beta^c, \alpha_1, \alpha_2, \alpha_3\}$, preference parameters $\{\psi^p, \psi^c, \delta_{i,t}, \Omega_\xi, \lambda_{i,t}, \tau\}$ can be estimated by simulated method of moments:

1. For a value of $\{\psi^p, \psi^c, \Omega_\xi\}$,

   (a) $\widehat{\delta}_{i,t}$ is computed based on equation (1.13).

$$\widehat{\delta}_{i,t} = \beta^p\Gamma^p_{t+1}\alpha_3\frac{Y_{i,t} - M_{i,t}}{M_{i,t}}$$

(b) $\{\widehat{\lambda}_{i,t}, \widehat{\tau}\}$ are computed based on equation (1.12). Suppose that parents i discuss performance in period t_1^i but not in period t_2^i, i.e., $S_{i,t_1^i} = 1$ and $S_{i,t_2^i} = 0$.

$$\lambda_{i,t_1^i}(1-\tau) = \beta^c \Gamma^c_{t+1} \alpha_2 \frac{\bar{L} - L_{i,t_1^i}}{L_{i,t_1^i}}$$
$$\lambda_{i,t_2^i} = \beta^c \Gamma^c_{t+1} \alpha_2 \frac{\bar{L} - L_{i,t_2^i}}{L_{i,t_2^i}}$$

$$\ln(1-\hat{\tau}) = \frac{1}{N^\tau} \sum_{i=1}^{N^\tau} \ln\left(\frac{\bar{L} - L_{i,t_1^i}}{L_{i,t_1^i}} \frac{L_{i,t_2^i}}{\bar{L} - L_{i,t_2^i}}\right)$$

2. Given the current values of all parameters, compute the following simulated moments:

   (a) mean and standard deviation of percent of income spending on educational investment in each period: $\mathbb{E}_t(\frac{M_{i,t}}{Y_{i,t}})$, $\sigma_t(\frac{M_{i,t}}{Y_{i,t}})$

   (b) mean and standard deviation of child effort in each period: $\mathbb{E}_t(L_{i,t})$, $\sigma_t(L_{i,t})$

   (c) fraction of parents who discuss performance with their children by demographics in each period: $\mathbb{E}_t(q_{i,t}|x_{1i}, x_{2i}, x_{3i})$

   (d) correlation between performance discussion and percent of income spending on educational investment in each period: $\rho_t(S_{i,t}, \frac{M_{i,t}}{Y_{i,t}})$

   (e) probability of switching performance discussion status: $\mathbb{P}(S_{i,t+1} = 1|S_{i,t} = 0)$, $\mathbb{P}(S_{i,t+1} = 0|S_{i,t} = 1)$

   (f) distribution of total numbers of performance discussion in T periods: $\mathbb{P}(J_i^S)$

   Let Λ_t^s be the simulated moment vector at period t.

$$\Lambda_t^s = [\mathbb{E}_t(\frac{M_{i,t}}{Y_{i,t}}), \sigma_t(\frac{M_{i,t}}{Y_{i,t}}), \mathbb{E}_t(L_{i,t}), \sigma_t(L_{i,t}), \mathbb{E}_t(q_{i,t}|x_{1i}, x_{2i}, x_{3i}), \rho_t(S_{i,t}, \frac{M_{i,t}}{Y_{i,t}})]'$$

   and $\Lambda^s = [\Lambda_1^s, \Lambda_2^s, ..., \Lambda_T^s, \mathbb{P}^s(J_i^S)]'$ be the simulated moment vector for all periods.

3. Find $\{\psi^p, \psi^c, \Omega_\xi\}$ that minimize the distance between simulated and data moments:

$$J = [\Lambda^s - \Lambda]' W [\Lambda^s - \Lambda] \tag{A.3}$$

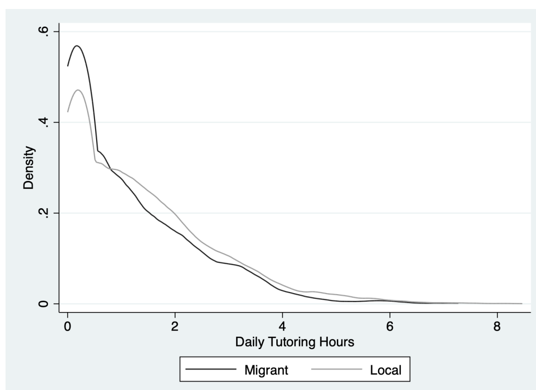
where W is a weighting matrix and $\Lambda = [\Lambda_1, \Lambda_2, ..., \Lambda_T, \mathbb{P}(J_i^S)]'$ are the corresponding data moments:

$$\Lambda_t = [\mathbb{E}_t(\frac{M_{i,t}}{Y_{i,t}}), \sigma_t(\frac{M_{i,t}}{Y_{i,t}}), \mathbb{E}_t(L_{i,t}), \sigma_t(L_{i,t}), \mathbb{E}_t(q_{i,t}|x_{1i}, x_{2i}, x_{3i}), \rho_t(S_{i,t}, \frac{M_{i,t}}{Y_{i,t}})]'$$

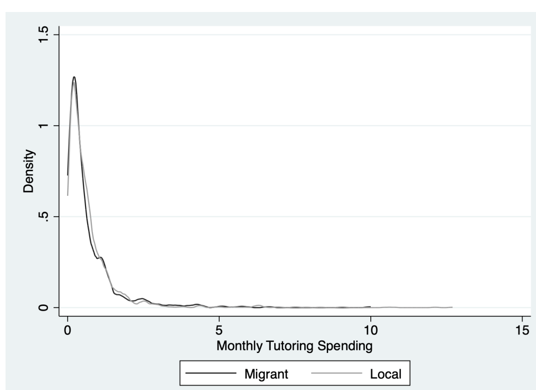
Appendix B

Appendix for Chapter 2

Figure B.1: The Distribution of Private Tutoring Investment



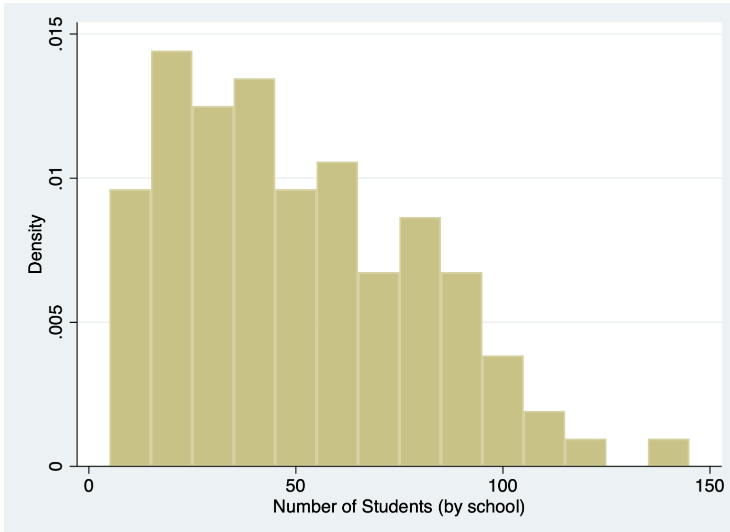
(a) Distribution of Daily Hours



(b) Distribution of Monthly Expense

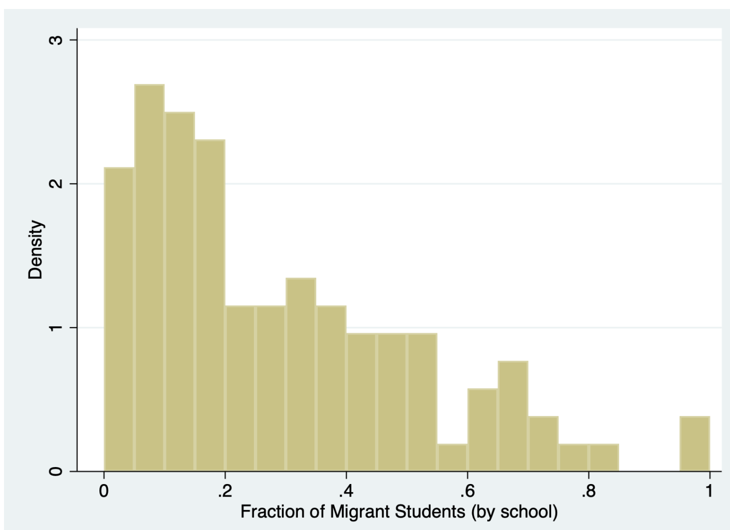
Source: CEPS.

Figure B.2: The Distribution of School Size



Source: CEPS.
Notes: This figure shows the empirical distribution of school size in the baseline sample.

Figure B.3: The Distribution of Migrant Student Fraction



Source: CEPS
Notes: This figure shows the empirical distribution of the fraction of migrant students in each school in the baseline sample.

Table B.1: Comparison of Baseline Sample and Full Sample

	(1) Baseline Sample		(2) Full Sample		Diff.
	Mean	SD	Mean	SD	(2)−(1)
Demographics					
Age	13.48	(0.64)	13.59	(0.73)	0.11***
Fraction of male students	0.51		0.53		0.01*
Fraction of migrants	0.29		0.20		-0.09***
Fraction of no sibling	0.58		0.43		-0.15***
Fraction of one sibling	0.31		0.42		0.11***
Fraction of two siblings	0.07		0.10		0.03***
Fraction of three/more	0.03		0.03		0.01***
Fraction of rich family	0.07		0.06		-0.01***
Fraction of middle income	0.79		0.72		-0.07***
Fraction of poor family	0.14		0.21		0.08***
Mother's years of schooling	10.83	(3.60)	9.68	(3.53)	-1.15***
Father's years of schooling	11.55	(3.29)	10.42	(3.15)	-1.13***
Test Scores					
Initial cognitive test score	0.14	(0.89)	0.00	(0.87)	-0.13***
New cognitive test score	0.41	(0.83)	0.30	(0.83)	-0.11***
Parental Investments					
Fraction taking tutoring	0.78		0.72		-0.06***
Conditioning on taking tutoring...					
- Daily tutoring hours	1.61	(1.26)	1.42	(1.21)	-0.19***
- Monthly expense (¥1K)	0.70	(1.01)	0.59	(0.96)	-0.10***
Good relation with mother	0.76		0.76		0.00
Good relation with father	0.65		0.66		-0.01
Fraction doing the following activities with child					
- Help with Homework	0.68		0.64		-0.04***
- Reading	0.69		0.65		-0.04***
- Watch TV	0.93		0.93		0.00
- Play sports	0.75		0.69		-0.06***
- Visit museum	0.85		0.78		-0.07***
- Watch shows	0.70		0.58		-0.12***
Educational Aspiration					
Going to college (child)	0.86		0.81		-0.05***
Going to college (parents)	0.85		0.80		-0.05***
No. of Observations	5,176		10,267		

Source: CEPS

Table B.2: Comparison of Selected Migrants and All Migrants

	(1) Public School Migrants		(2) All Migrants		Diff.
	Mean	SD	Mean	SD	(2)−(1)
Demographics					
Age	13.55	(0.67)	13.58	(0.69)	0.03
Fraction of male students	0.52		0.54		0.02
Fraction of no sibling	0.38		0.36		-0.02
Fraction of one sibling	0.45		0.47		0.02
Fraction of two siblings	0.12		0.12		0.00
Fraction of three or more	0.04		0.04		0.00
Fraction of rich family	0.06		0.06		0.00
Fraction of middle income	0.79		0.78		-0.01
Fraction of poor family	0.15		0.16		0.01
Mother's years of schooling	9.53	(3.36)	9.33	(3.32)	-0.20*
Father's years of schooling	10.48	(3.04)	10.27	(3.02)	-0.20*
Test Scores					
Initial cognitive test score	0.03	(0.87)	-0.02	(0.87)	-0.05
New cognitive test score	0.36	(0.81)	0.32	(0.82)	-0.04
Parental Investments					
Fraction taking tutoring	0.73		0.73		0.00
Conditioning on taking tutoring...					
- Daily tutoring hours	1.47	(1.18)	1.43	(1.16)	-0.04
- Monthly expense (¥1K)	0.68	(0.93)	0.66	(1.01)	-0.02
Good relation with mother	0.72		0.71		-0.01
Good relation with father	0.62		0.61		-0.01
Fraction doing the following activities with child					
- Help with Homework	0.64		0.63		-0.01
- Reading	0.67		0.65		-0.02
- Watch TV	0.94		0.94		0.00
- Play sports	0.72		0.70		-0.02
- Visit museum	0.84		0.82		-0.02
- Watch shows	0.65		0.62		-0.03*
Educational Aspiration					
Going to college (child)	0.83		0.80		-0.03*
Going to college (parents)	0.83		0.81		-0.02
No. of Observations	1,501		1,717		

Source: CEPS

Table B.3: The Effect of Private Tutoring on Cognitive Test Score

$\ln \theta_{i,1}$	(1)	(2)	(3)	(4)	(5)
Panel A: *OLS*					
Daily Tutoring Hours	0.0104**	0.00383	0.00145	-0.00641*	-0.0171***
	(0.00432)	(0.00357)	(0.00354)	(0.00358)	(0.00421)
Panel B: *IV*					
$\ln T_{i,0}$	0.0991***	0.0397***	0.0353***	0.00585	-0.0366**
	(0.0154)	(0.0125)	(0.0124)	(0.0147)	(0.0166)
No. of Students	5,176	5,176	5,176	5,176	5,176
Existing Skill		Yes	Yes	Yes	Yes
Demographics			Yes	Yes	Yes
Parenting Activities				Yes	Yes
School F.E.					Yes

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

Source: CEPS

Notes: In Panel A of Table B.3, I start with a simple regression of new cognitive test score on daily tutoring hours in column (1) and show how the relationship of private tutoring and cognitive test score changes, when different sets of controls are added.[a] The positive relationship becomes statistically insignificant when the existing cognitive skill, students' demographics, and parenting activities are controlled. When school fixed effects are added, the effect becomes statistically significantly negative. In Panel B of Table B.3, I present the results of the corresponding specifications of the production function technology where the tutoring investment is latent variable and measured with errors. The trend of the change in tutoring effect is similar, but the magnitude of the effect is larger in general.

[a]These are design-based regressions and variables are not latent or measured with errors.

Bibliography

Agostinelli, F., M. Saharkhiz, and M. Wiswall (2025). Home and school in the development of children. International Economic Review.

Agostinelli, F. and M. Wiswall (2016). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. Working paper, National Bureau of Economic Research.

Agostinelli, F. and M. Wiswall (2025). Estimating the technology of children's skill formation. Journal of Political Economy 133(3), 000–000.

Akabayashi, H. (1996). On the role of incentives in the formation of human capital in the family. Ph.D. dissertation, University of Chicago.

Andrabi, T., J. Das, A. Ijaz Khwaja, and T. Zajonc (2011). Do value-added estimates add value? Accounting for learning dynamics. American Economic Journal: Applied Economics 3(3), 29–54.

Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012). Estimating spillovers using panel data, with an application to the classroom. Quantitative Economics 3(3), 421–470.

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

Bhorkar, S. and M. Bray (2018). The expansion and roles of private tutoring in India: From supplementation to supplantation. International Journal of Educational Development 62, 148–156.

Bolt, U., E. French, J. Hentall-MacCuish, and C. O'Dea (2023). Intergenerational altruism and transfers of time and money: a life cycle perspective. Technical report, IFS Working Papers.

Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. Econometrica 83(3), 1147–1184.

Caucutt, E. M. and L. Lochner (2020). Early and late human capital investments, borrowing constraints, and the family. Journal of Political Economy 128(3), 1065–1147.

Chen, Y. and S. Feng (2013). Access to public schools and the education of migrant children in China. China Economic Review 26, 75–88.

Chen, Y. and S. Feng (2019). The education of migrant children in China's urban public elementary schools: Evidence from Shanghai. China Economic Review 54, 390–402.

Chen, Y. P. and Z. Liang (2012). Educational attainment of migrant children: The forgotten story of China's urbanization. In Education and Reform in China, pp. 117–132. Routledge.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. American Economic Review 104(9), 2593–2632.

Conley, T. G., N. Mehta, R. Stinebrickner, and T. Stinebrickner (2024). Social interactions, mechanisms, and equilibrium: Evidence from a model of study time and academic achievement. Journal of Political Economy 132(3), 824–866.

Cooper, H., J. C. Robinson, and E. A. Patall (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. Review of Educational Research 76(1), 1–62.

Cotton, C., B. Hickman, J. A. List, J. Price, and S. Roy (2020). Productivity versus motivation in adolescent human capital production: Evidence from a structurally-motivated field experiment. Working Paper 2020-150, University of Chicago, Becker Friedman Institute for Economics.

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Working paper, National Bureau of Economic Research.

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

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

Currie, J. and D. Almond (2011). Human capital development before age five. In Handbook of labor economics, Volume 4, pp. 1315–1486. Elsevier.

Del Boca, D., C. Flinn, E. Verriest, and M. Wiswall (2023). Parenting with patience: Parental incentives and child development. Working paper, CESifo.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. Review of Economic Studies 81(1), 137–185.

Del Boca, D., C. Monfardini, and C. Nicoletti (2017). Parental and child time investments and the cognitive development of adolescents. Journal of Labor Economics 35(2), 565–608.

Démurger, S., M. Gurgand, S. Li, and X. Yue (2009). Migrants as second-class workers in urban China? A decomposition analysis. Journal of Comparative Economics 37(4), 610–628.

Doepke, M. and F. Zilibotti (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. Econometrica 85(5), 1331–1371.

Doepke, M. and F. Zilibotti (2019). Love, money, and parenting: How economics explains the way we raise our kids. Princeton University Press.

Goodburn, C. (2009). Learning from migrant education: A case study of the schooling of rural migrant children in Beijing. International Journal of Educational Development 29(5), 495–504.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. Annu. Rev. Econ. 6(1), 689–733.

Hotz, V. J. and J. Pantano (2015). Strategic parenting, birth order, and school performance. Journal of Population Economics 28, 911–936.

Krueger, A. B. (1999). Experimental estimates of education production functions. The Quarterly Journal of Economics 114(2), 497–532.

Lee, S. Y. and A. Seshadri (2019). On the intergenerational transmission of economic status. Journal of Political Economy 127(2), 855–921.

Lu, Y. (2007). Educational status of temporary migrant children in China: Determinants and regional variations. Asian and Pacific Migration Journal 16(1), 29–55.

Lu, Y. and H. Zhou (2013). Academic achievement and loneliness of migrant children in China: School segregation and segmented assimilation. Comparative Education Review 57(1), 85–116.

Ning, J. (2021). Main data of the seventh national population census. China Stat 5, 4–5.

Ryu, D. and C. Kang (2013). Do Private Tutoring Expenditures Raise Academic Performance? Evidence from Middle School Students in South Korea. Asian Economic Journal 27(1), 59 – 83.

Todd, P. and K. I. Wolpin (2018). Accounting for mathematics performance of high school students in mexico: Estimating a coordination game in the classroom. Journal of Political Economy 126(6), 2608–2650.

Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. The Economic Journal 113(485), F3–F33.

Todd, P. E. and K. I. Wolpin (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. Journal of Human capital 1(1), 91–136.

Weinberg, B. A. (2001). An incentive model of the effect of parental income on children. Journal of Political Economy 109(2), 266–280.

Zhang, Y. (2013). Does private tutoring improve students' National College Entrance Exam performance?—A case study from Jinan, China. Economics of Education Review 32, 1–28.