ESSAYS IN LABOR ECONOMICS

by

Nicolás Badaracco

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Department of Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: 05/11/2021

The dissertation is approved by the following members of the Final Oral Committee:

Christopher Taber, Professor, Economics

Chao Fu, Professor, Economics

Jeffrey Smith, Professor, Economics

Matthew Wiswall, Professor, Economics

Jason Fletcher, Professor, Public Affairs and Sociology

A mi abuela Mausi por su cariño, apoyo, y sus milanesas.

Acknowledgments

I thank my advisor Chris Taber for his guidance and support throughout this project. His knowledge, brilliance, and patience were key inputs in this dissertation. I am grateful for his extremely reliable availability and good mood to talk about research. I want to thank my dissertation committee Chao Fu, Jeff Smith, and Matt Wiswall for providing comments and advice that significantly improve my dissertation. They provided me substantial amount their time and advice with great disposition. I have been very fortunate to have been advised and guided by them.

For additional comments, discussions, and advice, I want to thank Naoki Aizawa, Renata Gaineddenova, John Kennan, Rasmus Lentz, Arpita Patnaik, Hans Schwarz, Sandra Spirovska, Andrey Zubanov, and seminar participants at the University of Wisconsin-Madison.

This dissertation required significant amount of information that was collected and processed by several institutions. Without their work this dissertation would have not been possible. Chapter 1 and 2 use administrative data from the Agency for Quality of Education and from the Ministry of Education of Chile. I thank these institutions for allowing and facilitating access to this information. I am grateful to the Center for the Development of Inclusion Technologies (CEDETi UC) of the Pontifical Catholic University of Chile (PUC), to Marcelo Pizarro and his team for providing me results and explanations of the Chilean test WISC-V survey. Chapter 3 uses information from the Social Protection Survey (Ministerio Trabajo y Previsión Social, 2017). I thank the Undersecretariat of Social

Security for the authorization to use the Unnominated Database. All the results are my responsibility and not of any of these institutions and individuals.

Completar en el programa de doctorado fue posible gracias a mi familia. Agradezco a mis padres, Beatriz and Alberto, y mis hermanas y hermano, Ale, Tati, y Manu por su apoyo incondicional, por darme perspectiva, y su guía durante toda mi vida. Este agradamiento se extiende a toda mi familia en Burzaco y Eldorado. A mis abuelas Mausi y Nelly y mis abuelos Sonny y Tito. A mis tías y tíos, Moni, Cristi, Ame, Mariza, Marina, Guille, Carlos, Juani y Mario. A mis primas y primos, Dani, Lau, Maite, Ana, Mika, Tici, Maira, Mariano, Guido, Juani, Pablo y Pedro Felipe. También agradezco a mis amigos, a los mismos de siempre, Javier, José, Manuel, Alvaro, Pedro, y Leandro. Estar lejos de mis afectos ha sido sin ninguna duda el mayor costo en este emprendimiento.

Mi educación en Argentina fue fundamental para poder asistir al programa de doctorado en Madison. Agradezco enormemente la formación que me dio la Universidad Nacional de La Plata, tanto en la Licenciatura como en la Maestría en Economía. A su vez, mi formación en esta disciplina se beneficio por mi participación en el Centro de Estudios Distributivos, Laborales y Sociales (CEDLAS). Por eso, agradezco a Leonardo Gasparini y a todo el equipo del CEDLAS por la amistad, ayuda, y el aprendizaje.

Lastly, I want to thank Arpita Patnaik. We went through all this process together, side by side while sharing the ups and downs. Doing this together has been more rewarding, easier, and definitely more fun.

Abstract

The first chapter studies how households respond to school quality. I combine large-scale administrative and survey data from Chile to estimate parental and child time investment responses to classroom inputs and teachers from fourth to tenth grade. Since classroom inputs are not directly observable, I estimate a dynamic skill formation technology that provides classroom and teacher effects as a by-product, in a similar fashion as value-added models. I address selection by leveraging repeated observations of students and rich data on factors involved in household decisions. Parents of fourth graders compensate for low quality teachers and classroom inputs, while parents of high school students reinforce the quality of these inputs. Students, on the other hand, increase time self-investment if their classroom environment improves at every grade, but the responses are larger for older children.

The heterogeneous responses by grade found in Chapter 1 motivate the analysis of optimal resource allocation policies across education levels. Chapter 2 builds on Chapter 1 to understand how the differential impact by grade of school resources and home investment can be used to design the optimal allocation of the school resources across grades. To that end, I build and estimate a child development model using an indirect inference approach. I use the estimated model to simulate counterfactuals of the dynamics of the cognitive skills of students and I characterize the optimal allocation of school resources across grades. The results suggest that, on average, it is optimal to allocate relatively more resources in lower grades than in upper grades with respect to the allocation observed in the data.

Moreover, the behavioral response of households plays a key role in the characterization of the optimal allocation.

In the last chapter, I develop an empirical test for employer asymmetric learning about the productivity gains of On-the-Job (OTJ) training programs. I developed a model of OTJ training and employer learning. I solve the model under two types of learning: (i) asymmetric, current employer learns faster than potential employers and (ii) symmetric, the whole market learns simultaneously. The solution suggests different wage profile predictions under each form of learning. I build an empirical test based on these predictions and implement it on the Chilean Social Protection Survey by estimating a wage equation with interactions of training variables and tenure on the job. The results provide evidence of employer asymmetric learning.

Contents

List of Tables

List of Figures

Chapter 1

Time Investment Responses of Parents and Students to School Inputs

Chapter summary

The first chapter studies how households respond to school quality. I combine large-scale administrative and survey data from Chile to estimate parental and child time investment responses to classroom inputs and teachers from fourth to tenth grade. Since classroom inputs are not directly observable, I estimate a dynamic skill formation technology that provides classroom and teacher effects as a by-product, in a similar fashion as value-added models. I address selection by leveraging repeated observations of students and rich data on factors involved in household decisions. Parents of fourth graders compensate for low quality teachers and classroom inputs, while parents of high school students reinforce the quality of these inputs. Students, on the other hand, increase time self-investment if their classroom environment improves at every grade, but the responses are larger for older children.

1.1 Introduction

Cognitive development plays a key role in social and economic well-being. This process involves two main sources of inputs: home and school. The relationship between these inputs in the human capital accumulation process is complex. There are several agents providing investments, and it is unclear how the different inputs interact. Most of the existing literature has focused on home or school inputs in isolation.[1] A large body of work shows large teacher and classroom effects on educational achievement. If households make investments decisions based on their school environment, parents' and students' behavioral responses explain part of these estimates. For example, if parents compensate for low quality teachers, the value-added estimates for these teachers—the most extensively used measure of effectiveness—would be higher than in the absence of the parental response.[2] This is particularly relevant if the response varies by grade, since two equally good teachers assigned to different grades would then not have the same expected value-added estimate. This implies that behavioral responses to classroom inputs have implications on school resource allocation, teacher selection, and pay-for-performance policies.

In this chapter, I analyze the interaction of school and home investments in the child development process and the behavior of the actors that provide these investments. First, I study how parents and children adjust their time investments based on the quality of their school inputs. Second, I examine the evolution of the responses as children grow up. Third, given that extensive research shows a large contribution of teachers to educational achievement, I isolate the specific response of parents and students to teacher quality. Finally, motivated by the heterogeneity in responses by grade, I characterize the allocation

[1] The education production function literature has studied the effect of school inputs on academic achievement (see Hanushek (2020) for a survey). The child development literature in economics studies the skills formation process and households' investment decisions. See for example, Todd and Wolpin (2003, 2007), Cunha and Heckman (2008), Cunha et al. (2010), Fiorini and Keane (2014), Caetano et al. (2019), Del Boca et al. (2014), among others.

[2] In the education production function literature, teacher value-added represents the systematic variation in achievement across students assigned to the same teacher. See Hanushek and Rivkin (2012), Koedel et al. (2015), and Strøm and Falch (2020) for surveys on teacher value-added and related estimation methods.

of school resources across grades that maximizes the cognitive development of children.

To that aim, I use large-scale administrative data from Chile that provides information on the population of students and teachers and tracks them over time and across classrooms. This unique data reports standardized tests scores as well as parents' and children's answers to questionnaires on time investments—i.e., time parents spent with their children and time students spent on academic activities outside school—along with demographic characteristics. The data presents a challenge. The time investment questions have an ordered categorical structure and are not consistent across grades or calendar years.[3] To address this issue, I estimate each student's time investment measured in hours using a response model for these questions and the time investment distribution estimated from the Chilean Time Use Survey. Since classroom and teacher quality are not directly observed, I estimate the household's responses in two steps. First, I estimate the production function of cognitive skills of children which provides, as a by-product, measures of classroom and teacher quality. Second, I use these measures to estimate an approximation of the time investment policy function of parents and students.

To estimate the skill formation technology, I follow Agostinelli et al. (2020)'s framework that estimates classroom effects as the systematic variation in skills of students assigned to the same classroom. This methodology shares features with the child development literature, such as treating test scores as arbitrary scaled measures of latent cognitive skills and incorporating home inputs in the analysis.[4] In addition, my data allows me to estimate teacher effects as well. Since test scores in the data are not comparable across grades, identification of the dynamics of the skill formation process is challenging. To overcome this issue, I develop a measurement system of skills and I identify the dynamic system by exploiting additional survey data on cognitive development measures to track the evolution of the skills distribution across grades.

[3] Questions regarding time investments are not consistent across school grades or calendar years because they ask about different activities, the wording of the question changes, or the possible answers change.

[4] Work in the child development literature shows that ignoring mis-measured skills can lead to substantial bias in the estimation of the skill formation technology (Cunha and Heckman, 2008; Cunha et al., 2010).

I then use the measures of classroom and teacher quality to estimate the time investment responses of households. There are two threats to identification: 1) estimation error in the classroom and teacher effects and 2) potential selection on unobservables. To tackle the first issue, I estimate the responses to classroom and teacher effects using a two-stage least squares (2SLS) estimator. To acquire additional measures of classroom and teacher effects, I estimate the skill technology multiple times by randomly selecting half of the students in each classroom—in the spirit of *leave-one-out* estimators. Each iteration provides additional measures that provide exclusion restrictions to identify the responses. The second concern implies that the empirical correlation between classroom quality and time investment could be attributed to the assignment of students across classrooms based on unobservables. To address this, I leverage the multiple observations per student to control for time-invariant unobservables and rich data on the relevant factors involved in the household's decision. The resulting procedure estimates an approximation of the household's time investment policy function.

The estimates of the time investment responses are not homogeneous across grades. Consider reassigning a student from the 25th to the 75th percentile of the classroom quality distribution. For students in grade 4, parents compensate by decreasing parental time by around 1.8 weekly hours. The magnitude decreases as children grow up. For tenth graders, in contrast, parents reinforce classroom quality by increasing parental time by 45 minutes per week. Students, on the other hand, reinforce classroom quality at all ages—between 20 and 30 minutes per week—and responses are larger for older children. These responses are quantitatively significant, representing over 10 percent of the average time investment. Furthermore, the households' responses to classroom quality imply a non-trivial effect on the cognitive skills of children, representing between 3 and 11 percent of the total effect of classroom quality (depending on the school grade). Meanwhile, household's responses to teacher quality follow a similar pattern with smaller magnitudes, although students at grade 10 are unresponsive.

This chapter contributes to two large bodies of research that study the human capital accumulation process of children. First, the education production function literature that studies the influence of school inputs in academic performance of students, and second, the child development literature that focuses on home investments in the development process. The contributions build bridges between these two strands of the literature but some are particularly more relevant for one of these blocks of research.

First, I contribute to both branches of the literature by analyzing a more complete picture of the technology of cognitive skills formation of children. Some studies, such as Todd and Wolpin (2007) and Agostinelli et al. (2020), consider both home and school inputs in the development process. I build on their work by also incorporating in the technology home investment provided by different members of the household (parents and children) and specifications with teachers' and other observable classroom inputs' contribution—as opposed to a unique generic home and school investment measure. Moreover, I consider a dynamic skill formation technology across school grades that allows for the effects of inputs and the relationship between inputs to be grade-specific. The additional dimensions of inputs and the flexibility of the technology across grade improve our understanding of the inputs' effects and the relationship between inputs in the skill formation process.

I also contribute by studying the behavior of the actors providing investments. There is a small but growing line of research studying the responses of families to school inputs.[5] These papers study responses to specific classroom components or proxies of school quality.[6] Their results can inform the design of specific policies. For example, family responses to peer composition inform the design of student tracking policies but are less helpful for the analysis of teacher-related policies. The magnitude, and even the direction, of

[5]See Rabe (2020) for a survey of studies on families' responses to school inputs.

[6]For example, class size (Datar and Mason, 2008; Fredriksson et al., 2016), per-pupil expenditure (Houtenville and Conway, 2008), grants (Das et al., 2013), enrollment in preferred school (Pop-Eleches and Urquiola, 2013), peer composition (Fu and Mehta, 2018; Agostinelli, 2018), elicited parental beliefs on school quality (Attanasio et al., 2018), school quality information (Greaves et al., 2019), teachers' qualifications and training programs (Chang et al., 2020; Gensowski et al., 2020), among others. Meanwhile, Nicoletti and Tonei (2020) and Jacqz (2020) study parental responses to human capital shocks and Berniell and Estrada (2020) analyze parental responses to age of school entry.

the responses to specific inputs might be different than to the combination of all inputs provided by the school—or teachers in particular.

In addition, I consider heterogeneity in responses by school grade and by members of the household that provide the investment. Stylized facts from existing work show that home investments, such as parental and child time investment, vary substantially as children grow up, and while children's investment increases with age, it decreases for parents (Del Boca et al., 2014). Furthermore, the literature shows that these investments have a differential impact on cognitive skills by age.[7] These results suggest a dynamic skill formation technology and different investment costs across school grades, which potentially could lead to heterogeneity in responses of parents and children by grade. Thus, I estimate grade-specific responses of both parents and children between grades 4 to 10, in contrast with existing work that do not compare responses for children of different age and mainly focus on parental investments.[8] The estimates of the parents' and children's responses to classroom and teacher quality across grades inform about the role of households and their interaction with schools in the development process of children.

In particular, the estimates of the households' responses contribute to the education production function literature by improving our understanding of teachers' and classrooms' influence in the development of children. This literature interprets the estimates of teacher and classroom value-added as their intrinsic quality. The estimates of the household responses shed light on the mechanics in play within the black box of classroom and teacher value-added and quantify the behavioral component in these measures of effectiveness. This information can be used in the design of school resource allocation and teacher-related policies.

The structure of the chapter is as follows. Section 1.2 develops a child development

[7] Carneiro and Heckman (2003) and Del Boca et al. (2014) show parental time's effect diminishes as children grow while Cooper et al. (2006) and Del Boca et al. (2019) show that the effect of child time investment increases.

[8] An exception is Greaves et al. (2019) that finds that children increase effort in light of positive information about school quality.

model and specifies the household problem, the skill formation technology and time investments functions. Section 1.3 presents the data and descriptive statistics. Section 1.4 describes the measurement systems of skills and time investment. Sections 1.5, 1.6, and 1.7 present identification, estimation methodology and estimates, respectively. Each section addresses measurement systems, skill technology, and time investments functions. Finally, Section 2.4 concludes the chapter.

1.2 Child development model

I build a child development model which follows existing models in the literature, such as Del Boca et al. (2014, 2016, 2019), Caucutt and Lochner (2020), and Agostinelli (2018), among others. The key difference with existing work is that I explicitly incorporate classroom inputs as a composite input which summarizes all school inputs in a particular classroom–i.e., teacher quality, peer composition, money resources, etc.[9]

1.2.1 Household problem

The model consists of a unitary household comprised of parents and a single child. Throughout the chapter the index i is used interchangeably between child and household and the index t between the age of the child and the school grade she attends. The household maximizes its lifetime utility by choosing the amount of time to invest in the cognitive skill formation of the child at each school grade.[10] Let $\Omega_{it} = \{\theta_{it}, C_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\}$ be the state space, where θ_{it} is the child's cognitive skill, C_{it} is classroom inputs, and the vectors $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ are exogenous characteristics. The vectors $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ might share elements and they influence preferences and the skill formation technology, respectively. The instantaneous utility

[9] In this model, I do not consider explicitly the interactions between parents and children (De Fraja et al., 2010; Caetano et al., 2019; Del Boca et al., 2019). Albornoz et al. (2018) develop a theoretical model that analyzes the interaction between student, parents and teachers.

[10] The household problem in a more general setting could include additional decisions, such as consumption and leisure. However, the data has no information on hourly wages, working time, or leisure. Thus, I specify the problem only in terms of the skills accumulation and time investments.

function is denoted by u_{it} and its arguments are the child's skill stock at the beginning of the period θ_{it}, time investment h_{it} (vector size D), and the vector $\mathbf{x}_{it}$. The value function of the recursive problem is denoted by $V_{it}(\cdot)$. Both utility and value functions are indexed by i allowing for an idiosyncratic component. Lastly, H is the household time endowment at each period and β is the discount factor. To simplify notation, I omit the i index in this section. The problem of the household is:

$$\begin{aligned} V_t(\Omega_t) \equiv \max_{h_t} & \left\{ u_t(\theta_t, h_t, \mathbf{x}_t) + \beta \mathrm{E}[V_{t+1}(\Omega_{t+1})] \right\} \\ \text{subject to} & \\ & \text{Time constraint} \qquad h_t \in [0, H]^D; \\ & \text{Production function} \quad \theta_{t+1} = F_t(\theta_t, h_t, C_t, \mathbf{z}_t, \nu_t), \end{aligned} \tag{1.2.1}$$

where the control variable h_t is the time the household invests in the child. I assume that, conditional on Ω_t, the household does not choose a classroom—i.e., classroom inputs are exogenous.[11] The skills' dynamics is governed by the technology defined by $F_t(\cdot)$, which is a function of current skill, the time the household invests in the child, classroom inputs, other observable characteristics $\mathbf{z}_t$, and an unobserved (to the household and econometrician) shock ν_t. Note that this function is indexed by t allowing for the effects of inputs to be specific to the grade the child is attending. The household knows the skill formation technology $F_t(\cdot)$ and it has rational expectations. The household forms expectation over the skill shock ν_t and future classroom inputs.

Note that time investment h_t can be multidimensional. For example, household members can provide time investments with different implications in the cognitive development of the child. In the empirical implementation there are two time investment choices ($D = 2$), parental and child time investment. The former is the time parents spend with their chil-

[11] A more general model can incorporate school choice and allow this decision to affect the expected classroom input at each school grade. However, after that choice is made, the actual peer composition or the teacher in the classroom is out of the control of the household. Additionally, this requires the school choice to be irreversible; at least in terms of the expected quality of the school inputs. Thus, classroom inputs exogeneity, conditional on Ω_t, is arguably a weak assumption.

dren while the latter is the time children spend (on their own) studying, doing homework, or engaging in other academic activity while not in school. However, without loss of generality, I lay out the model, the identification and estimation methodologies assuming time investment is one-dimensional—i.e., $D = 1$—in order to simplify notation and facilitate tractability. Nevertheless, generalization to higher dimensions is straightforward.

With one-dimensional investment, the first order condition (of the interior solution) is given by the following equation:

$$\frac{\partial u_t(h_t, \theta_t, \mathbf{x}_t)}{\partial h_t} + \beta \frac{\partial \mathrm{E}[V_{t+1}(\Omega_{t+1})]}{\partial h_t} = 0. \tag{1.2.2}$$

The first term of equation (1.2.2) is the marginal disutility cost of the time investment, and the second term is the marginal benefit, given by the change in the expected continuation value. The optimal level of investment that solves equation (1.2.2)—i.e., the policy function—can be expressed as:

$$h_t \;=\; h_t^f(\theta_t, C_t, \mathbf{x}_t, \mathbf{z}_t). \tag{1.2.3}$$

If the marginal disutility cost does not depend on classroom inputs, then the relationship between time investment and classroom inputs is determined by how the classroom environment affects the marginal benefit of time investment. If a higher quality of school inputs increases the marginal benefit, the household responds by investing more in the child and *vice versa*. Ignoring the expectation operator (i.e., assuming no uncertainty) for the sake of exposition, the partial derivative of the marginal benefit with respect to classroom inputs is:

$$\frac{\partial^2 V_{t+1}(\Omega_{t+1})}{\partial \theta_{t+1}^2} \frac{\partial F_t}{\partial C_t} \frac{\partial F_t}{\partial h_t} + \frac{\partial V_{t+1}(\Omega_{t+1})}{\partial \theta_{t+1}} \frac{\partial^2 F_t}{\partial h_t \partial C_t}. \tag{1.2.4}$$

The interpretation of equation (1.2.4) is straightforward. The first term accounts for the change in the marginal benefits as a result of the curvature of the value function weighted by the productivity of both inputs. For example, if the value function is concave, a higher skill stock as a result of an improvement in school inputs decreases the marginal utility of time investment, so this term is negative. Assuming skills increase the continuation value, the second term's sign depends on the relationship between these two inputs in the technology of skill formation. If $\partial^2 F_t/\partial h_t \partial\theta_t < 0$, these inputs are substitutes in production, so that higher levels of classroom quality are associated with lower productivity of time investment. Instead, $\partial^2 F_t/\partial h_t \partial\theta_t > 0$ implies complementarity in production, and classroom quality makes the households investment more productive.[12] The overall sign depends on preferences and on the skill technology; ultimately, this is an empirical question.[13]

1.2.2 Skill formation technology

This section describes the dynamics of the skill accumulation process. The cognitive skills of a child θ_{it} follow a first order Markov process. This is consistent with the specifications of the skill formation technology in Cunha and Heckman (2008), Cunha et al. (2010), Agostinelli and Wiswall (2016), and Agostinelli et al. (2020). The skill formation process is:

$$\theta_{it+1} = F_t(\theta_{it}, h_{it}, C_{it}, \mathbf{z}_{it}, \nu_{it}), \tag{1.2.5}$$

where $F_t(\cdot)$ is a grade-specific function that depends on current skill, θ_{it}, the time the household invest in its child h_{it}, the classroom inputs C_{it}, household characteristics $\mathbf{z}_{it}$, and the structural shock ν_{it}.

[12]See Cunha et al. (2006) for a discussion on the definitions of complemetarity and substitutability in the skill formation technology.

[13]In the multidimensional case of time investments, the direction of the response is ambiguous as well. However, the analysis is more complex and it depends on the relative productivity and complementarity/-substitutability between different time investments.

Baseline parametrization

I assume the technology is a trans-log production function. However, the relationship between the logarithms of skills and time investment is linear, allowing for a corner solution.[14] This parametrization is flexible in terms of the relationship between inputs in the production of skills—i.e., it is possible to have a negative cross-derivative (see Agostinelli and Wiswall, 2016). It allows complementarity or substitutability between time and classroom inputs.[15] The parametric functional form is the following:

$$\begin{aligned}\log \theta_{it+1} &= \log F_t(\theta_{it}, h_{it}, C_{it}, \mathbf{z}_{it}, \nu_{it}) \\ &= \log A_t + \gamma_{1t}\log\theta_{it} + \gamma_{2t}h_{it} + \gamma_{3t}\log C_{it} + \gamma_{4t}h_{it}\times\log C_{it} + \mathbf{z}_{it}'\gamma_{5t} + \nu_{it},\end{aligned} \tag{1.2.6}$$

where $A_t \exp(\mathbf{z}_{it}'\gamma_{5t})$ is the total factor productivity. The set $\{\gamma_{jt}\}_{j=1}^{5}$ defines the elasticity or semi-elasticity of next grade's skills and inputs and ν_{it} is a mean zero shock.

The specification is more general in the empirical implementation. Besides the terms in equation (1.2.6), it includes second order polynomials of current skill and time inputs, interactions between time investment, classroom inputs, and current skills and interactions between time investments of different members of the household—i.e., parental and child time investments. However, equation (1.2.6) reduces notation burden and the identification and estimation analysis under this simplification is without loss of generality.

Within-classroom components

The previous setting includes the effect of all observable and unobservable (to the econometrician) classroom components as a composite input. The education production function literature studies the effects of different classroom inputs on students' performance (see

[14]In the data, I observe a non-trivial fraction of households choosing to invest zero time in the skill formation process.

[15]Assuming a Cobb-Douglas or constant returns to scale production function with standard parameters values implies weakly complementarity between all the inputs. It is important to allow for this flexibility since the signs of cross-derivatives are relevant in terms of households' responses.

Hanushek, 2020). In particular, there is an extensive line of research studying the contribution of teachers to the academic achievement of students. These results motivate the analysis of the response of households to teacher quality. I use an additional specification that decomposes the classroom inputs into its different components. This allows for a more flexible technology in terms of the complementarity or substitutability between teacher quality and time inputs. Teacher effects are denoted by $\log T_{it}$ while $\mathbf{r}_{it}$ and ξ_{it} are the observed and unobserved (to the econometrician) classroom inputs, respectively.[16] The technology of the classroom inputs is:

$$\log C_{it} = \psi_{1t} \log T_{it} + \mathbf{r}'_{it}\psi_{2t} + \xi_{it}, \tag{1.2.7}$$

where ψ_{1t} and ψ_{2t} represent each of these components' contributions to the classroom effects. Under this specification, the skill formation technology has the following structure:

$$\begin{aligned} \log \theta_{it+1} = \; & \log A_t + \gamma_{1t} \log \theta_{it} + \gamma_{2t} h_{it} + \gamma_{3t} \log T_{it} + \mathbf{r}'_{it}\gamma_{4t} \\ & + \gamma_{6t} h_{it} \times \log T_{it} + h_{it} \times \mathbf{r}'_{it}\gamma_{7t} + \mathbf{z}'_{it}\gamma_{8t} + \nu_{it}. \end{aligned} \tag{1.2.8}$$

It should be noted that the error term ν^T_{it} includes unobserved classroom level inputs ξ_{it}—in contrast with ν_{it}in equation (1.2.6). Thus, identification of these parameters requires stronger assumptions than the ones in equation (1.2.6).

1.2.3 Time investment function

The amount of time households decide to invest in their children is an equilibrium object, since it is the solution to the household problem. The policy function of time investment is denoted by $h_{it} = h^f_{it}(\theta_{it}, C_{it}, \mathbf{x}_{it}, \mathbf{z}_{it})$. It is a function of current skill, classroom inputs and

[16] In the empirical implementation the observed classroom inputs are classroom average share of male students, household income, parents' age, child effort and parental time, shares of parents' education levels, share of poor students, class size, number of teachers and subjects, average skill of peers, teacher-student gender match indicator, and share of classmates in the bottom and top 5 percent of the skill distribution.

observable inputs. The policy function might have an idiosyncratic component—induced by heterogeneous preference—that leads to different time allocation choices.[17] The household response to classroom quality is then given by the partial derivative of h^f_{it} with respect to C_{it}. Following Cunha et al. (2010), Agostinelli and Wiswall (2016), Attanasio et al. (2020b,a), among others, I use a parametric specification that represents an approximation of the (unknown) true underlying policy function:

$$h_{it} = \delta_{0,t} + \delta_{C,t} \log C_{it} + \delta_{\theta,t} \log \theta_{it} + \mathbf{\Gamma}'_{it}\delta_{\mathbf{\Gamma},t} + \pi_i + \eta_{it}, \tag{1.2.9}$$

where $\{\delta_0, \delta_{C,t}, \delta_{\theta,t}, \delta_{\mathbf{\Gamma},t}\}_t$ are parameters, $\mathbf{\Gamma}_{it} \subseteq \{\mathbf{x}_{it}, \mathbf{z}_{it}\}$ is a vector of time-varying demographic characteristics, π_i is a household idiosyncratic component and η_{it} is a disturbance term. Under the specification that includes within-classroom components, the time investment function is:

$$h_{it} = \delta_{0,t} + \delta_{T,t} \log T_{it} + \delta_{\theta,t} \log \theta_{it} + \mathbf{r}'_{it}\delta_{r,t} + \mathbf{\Gamma}'_{it}\delta_{\mathbf{\Gamma},t} + \pi_i + \eta_{it}. \tag{1.2.10}$$

In equation (1.2.10) I allow for household responses to teachers $(\log T_{it})$ to be different than responses to other observable classroom inputs $(\mathbf{r}_{it})$. Similar to the case of the skill technology, the required identification assumptions over η_{it} are stronger, since there are additional parametric assumptions on the technology of classroom inputs and potential unobserved classroom inputs.[18] If equations (1.2.9) and (1.2.10) incorporate all relevant factors influencing the household decision, these equations are an approximation of the policy function of the time investment. Making the policy function linear in the logarithm of classroom inputs and skills follows from the parametric assumption of the skill technology. Nevertheless, the interpretation of the coefficients is relative to movements across the

[17] An idiosyncratic component of the utility function implies that the function $h^f_{it}(\cdot)$ is indexed by i.

[18] The household should care about all teachers' characteristics that contribute to the skill formation. Thus, I redefined teacher effects in equation (1.2.10) to be the innate teacher effect, $\log \overline{T}_{it}$, and the contribution of tenure at the school and teaching experience—i.e., $\log T_{it} = \log \overline{T}_{it} + \text{pol}^{\text{ten}}_t(\text{tenure}) + \text{pol}^{\text{exp}}_t(\text{experience})$, where pol_t is a second order polynomial function.

distribution of classroom inputs—e.g., moving a student from the 25th to 75th percentile in the classroom quality distribution.

Estimating equations (1.2.9) and (1.2.10) provides estimates of the responses of households without the need to specify and estimate the full household model in equation (1.2.1). However, this equation can only evaluate policies that change resources in a particular school grade, holding everything else constant—i.e., expectations of future classroom environments. To evaluate a wider set of policies, it is necessary to specify and estimate the dynamic decision process. Del Boca et al. (2014, 2019) and Cunha et al. (2013) explicitly model household preferences and beliefs and estimate models which allows them to evaluate counterfactual policies that require household to update expectations. In this chapter, I follow both approaches. I estimate the reduced form (in the literal sense) given by equations (1.2.9) and (1.2.10) that provide the household responses. Then, with additional structure on preferences and the expectation process, I estimate the child development model from equation (1.2.1), which allows me to evaluate a broader policy space.

1.3 Data and descriptive evidence

1.3.1 Data

In this chapter, I use large-scale administrative data of Chile from two sources. These databases are called *Sistema de Información General de Educación* (SIGE) and *Sistema de Medición de la Calidad de la Educación* (SIMCE). The SIGE database is provided by the Ministry of Education of Chile and it contains information on the entire education system of Chile. That is, this database has information on every student, teacher, and school, and it tracks students and teachers over time and across schools, classrooms, and subjects. It reports basic demographic information of students, such as age and gender, as well as academic information—e.g., students' subject grades, GPA, attendance, among others. In addition, it provides teachers' characteristics, such as age, teaching experience, tenure at school, and education.

The SIMCE database is provided by the Agency for Quality of Education.[19] It has scores from standardized tests designed to measure academic achievement. The objective of this exam is to evaluate the performance of schools and track its evolution. Since the late 1990s, every year students in specific grades take these exams in a set of subjects, close to the end of the academic year (October/November).[20] Table S.1.1 indicates for which grades and years there is information available for the time period relevant for the analysis.[21] For every year there is a math and language exam plus at least one additional exam in a particular subject—e.g., natural sciences and social sciences. These exams are meant to evaluate the curricula established by law at each grade in Chile.

The test scores from these exams are generated using item response theory. Under certain conditions, described in Ballou (2009), the test scores are interval scales—i.e., each point represents the same amount of learning or skill at any point of the distribution.[22] This property implies that the test score scale is invariant to affine transformations, but not invariant to monotone transformations. However, the test scores are not comparable across school grades and I explain how I deal with this issue in the following sections. Either way, it should be noted that many authors have suggested that test scores should be treated as ordinal measures, as opposed to interval scaled (Ballou, 2009; Jacob and Rothstein, 2016). We should be cautious when drawing conclusions from the results that depend on test scores being treated as interval scaled, as opposed to ordinal. For example, the interpretation of the household responses does not depend on the scale of the test scores, but the scale's properties play a key role in comparing the effects of inputs on cognitive

[19]For details, see the technical report Agencia de Calidad de la Educación (2015).

[20]One drawback of the database is that the exams are not taken in consecutive grades.

[21]The time frame of the analysis is between 2011 and 2018. During this period the exams were taken by students in 2nd, 4th, 6th, 8th and 10th grade. Since in 2008 there was a large voucher program that change significantly the school market (Neilson, 2013), I use information post 2011 to avoid variation related to this policy. Moreover, the first year the exams took place in second grade was 2012.

[22]A measure is interval scale if the ratio between two intervals is unit free—i.e., the measure has two degrees of freedom: unit and origin. Ballou (2009) and Jacob and Rothstein (2016) point out the importance of this property to estimate value-added models.

skills across grades.

A unique feature of the SIMCE data is that while these exams take place, every parent and student taking the test fill in a questionnaire (separately). These consist of a series of questions about household characteristics, such as household income, parents' education and age and information related to the time parents spent with their child and the time the student spent studying, doing homework or other academic activities. Hanushek and Rivkin (2012) point out that it is very unusual to have information on the students performance together with home investments measures, making the SIMCE data well suited to study the relationship between classroom inputs and the time investment of households.[23]

I use two additional sources of survey data. One is the Time Use Survey of Chile, which is a nationally representative household survey collected in 2015. The survey collects information on the members of the sampled households regarding the time they spent in activities taking place in the most recent weekday and weekend day. In particular, it provides information on parents' and children's time allocations: hours parents spent with their children and time children spent on academic activities outside school.

The second survey was collected by the Center for the Development of Inclusion Technologies (CEDETi UC) of the School of Psychology of the Pontifical Catholic University of Chile (PUC). This survey was collected in 2017 and it is a nationally representative sample of the population of children between 6 and 16 years old. The data reports test scores on the Wechsler Intelligence Scale for Children, fifth edition (WISC-V). It consists of a set of fifteen cognitive tests that aim to evaluate the cognitive development of children. The CEDETi UC collected this survey in coordination with government institutions to define the standards of the WISC-V test for the population of Chile.

It should be noted that the administrative data is virtually a census of the population

[23] Bharadwaj et al. (2018) use data from the same source to explore the relationship between health at birth, academic outcomes and the role of parental investments.

of children at each age while both surveys are representative of the same population.[24] Appendix A describes the data in additional detail.

1.3.2 Descriptive statistics

In Chile, children start attending primary school when they are 6 years old. Primary school consists of grades 1 to 8. By the end of second grade most students are 8 years old—i.e., when the younger children in the sample take their first SIMCE exam.[25] Secondary school includes grades 9 to 12. At the end of tenth grade, students take their last SIMCE exam, when most of them are 16 years old.

I drop all students in classrooms with less than 10 students or with missing test scores or time investments measures. Table 1.1 shows descriptive statistics of students. The sample is an unbalanced panel and the number of students at each school grade varies. Around 80 percent of the parents' questionnaires were answered by the students' mothers. On average, they are 38 years old for fourth graders and almost 44 years old for students in grade 10. I split the education level of parents in three categories: less than high school, high school, and more than high school. The students in the analysis sample are from slightly more affluent households than the entire population. The education distribution of fathers is roughly a third at each education category while mothers' distribution is skewed towards higher education levels. Monthly household income is around 600 thousand Chilean pesos, equivalent to around 900 dollars (in 2018 values). Classroom size is large relative to the US and other developed countries, with an average class size between 33 and 36 students. Similar to most education systems, the average number of teachers interacting with students and subjects increases in higher levels of education, but relatively less for the latter.

[24]The dropout rate is virtually zero for primary school in Chile and below 5 percent in grade 10. This should not lead to any substantial issues in terms of representativeness between the administrative and survey data.

[25]In Chile the cutoff date to start formal education is March and the academic year overlaps the calendar year. That is, all children who turn 6 years old by March of the academic year attend first grade. By the end of the academic year around 75 percent have turned 7 years old.

Table 1.1:
Descriptive statistics - Students

	Fourth grade		Sixth grade		Eighth grade		Tenth grade	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female student (%)	50.6		51.5		50.5		49.8	
Mom answered (%)	79.8		81.5		80.4		78.9	
Age parent (years)	38.3	7.3	40.1	7.2	42.0	7.1	43.9	7.0
Father <HS (%)	28.3		29.2		32.6		30.8	
Father HS (%)	36.4		36.6		35.5		36.4	
Father >HS (%)	35.2		34.2		31.9		32.8	
Mother <HS (%)	25.1		26.4		30.0		27.8	
Mother HS (%)	38.4		38.6		38.0		39.3	
Mother >HS (%)	36.5		35.0		32.0		33.0	
Monthly household income (ths. CLP)	617.2	579.9	608.5	575.7	581.8	557.6	634.5	567.3
Class size	33.4	7.6	34.7	7.4	34.0	7.2	36.5	6.6
No. teachers	5.8	2.2	8.6	1.4	8.5	1.2	10.5	1.4
No. subjects	9.8	0.8	10.3	0.6	9.5	0.8	11.3	1.4
Classrooms	19,370		30,188		24,568		18,567	
Students	407,720		596,617		457,782		336,470	

Note: Sample consists of students in classrooms with at least 10 students with non-missing values in test scores and time investment questions. Mom answered is an indicator that the mother answered the questionnaire. Household income is in thousands Chilean pesos in 2018 values ($1 dollar∼ $650 Chilean pesos). HS refers to high school education.

The estimates involving teachers require additional sample restrictions.[26] I drop students whose math teachers I do not observe in at least two years in their careers.[27] Table 1.2 shows the characteristics of math teachers in this sample. They are on average around 40 years old at every school grade.[28] The share of female teachers decreases across grades from more than 80 percent in grade 4 to 52 percent in grade 10. Additionally, teachers in lower grades tend to have proportionally more education degrees relative to other degrees, and they have more teaching experience and higher tenure at the school than teachers in

[26]The sample restrictions are required to identify teacher effects and are usual in the literature to estimate teacher value-added.

[27]I consider math teachers since I use math test scores to estimate teacher effects. I make this restriction to separately identify the effects of teaching experience and tenure at school from the time-invariant teacher effects—i.e., I need to observe teachers in at least at two different points in their careers.

[28]See Behrman et al. (2016) and Tincani (2020) for work that studies the market of teachers in Chile.

Table 1.2:
Descriptive statistics - Teachers

	School grade							
	Fourth		Sixth		Eighth		Tenth	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female (%)	83.6		66.1		59.4		52.1	
Age (years)	41.1	11.1	40.3	11.9	40.9	12.0	40.2	12.2
Education degree (%)	97.2		94.7		93.9		89.8	
Other degree (%)	2.8		5.3		6.1		10.2	
Teaching experience (years)	14.1	11.4	12.7	12.1	13.1	12.1	12.6	11.6
Tenure at school (years)	9.2	9.0	7.6	9.3	7.7	9.2	7.7	9.0
Teachers	3,103		5,683		4,728		3,462	

Note: Sample consists of math teachers in classrooms with at least 10 students with non-missing values in test scores and time investment questions. I include teachers assigned to at least two classrooms in different calendar years.

upper grades.

Table 1.3 shows descriptive statistics of parental and child time self-investment. As reported in existing work (e.g., Del Boca et al., 2014), parents decrease the time they spend with their children as they grow up. Average parental time is 14.5 hours per week when children are 10 years old and 6.5 hours per week for 16 year olds. Children increase time self-investment as they grow up—from 2 to almost 7 hours per week. Figure S.1.1 shows the distribution for both time investments by school grade.

1.3.3 Time investments and classroom inputs

The SIMCE questionnaires answered by parents and students include a series of questions regarding parental and child time investment. These are framed as ordered categorical questions—e.g., *How often do you help your child with her homework? Answers: never, sometimes, often, always*.[29] Some of these questions vary across calendar years and school grades because wording of the question changes, the possible answers change or they ask about different activities. Thus, direct comparisons between these variables across school grades,

[29] Table S.1.2 reports a subset of questions and their structure translated to English.

Table 1.3:
Descriptive statistics - Time investments

Child's age	School grade	Parental time (weekly hours)		Child time (weekly hours)	
		Mean	SD	Mean	SD
10	4th	14.5	13.1	1.9	1.9
12	6th	13.0	12.8	4.5	3.8
14	8th	9.6	12.1	4.8	3.2
16	10th	6.5	8.6	7.6	6.9

Note: Calculations are based on the Chilean Time Use Survey 2015. The survey reports hours spent in activities in the last weekday and weekend day. Parental time refers to hours parents spend in activities with their children and child time is hours a child spends studying, doing homework or other academic activities outside school.

or even within grade but across calendar years, is not possible.

Using only information from the administrative data, I perform an exercise that allows me to analyze the *direction* of the household response to classroom inputs. First, I estimate a classroom value-added model for each school grade. It identifies and estimates classroom value-added on student achievement.[30] Second, I estimate a regression of each time investment question on the classroom value-added estimates. The sign of the coefficients provide evidence of the direction of the response to school inputs.[31]

Figure 1.1 plots the coefficients of the regressions. Each dot represents the coefficient using a different time investment question (standardized to have mean zero and standard deviation one). The units on the vertical axis represent the response as a percent of the standard deviation of the time investment question. These numbers are not comparable

[30] Let A_{ijt} be the academic achievement of student i at classroom j in school grade t, measured by the math test score standardized at the school-grade level. The classroom value-added model has the following specification:

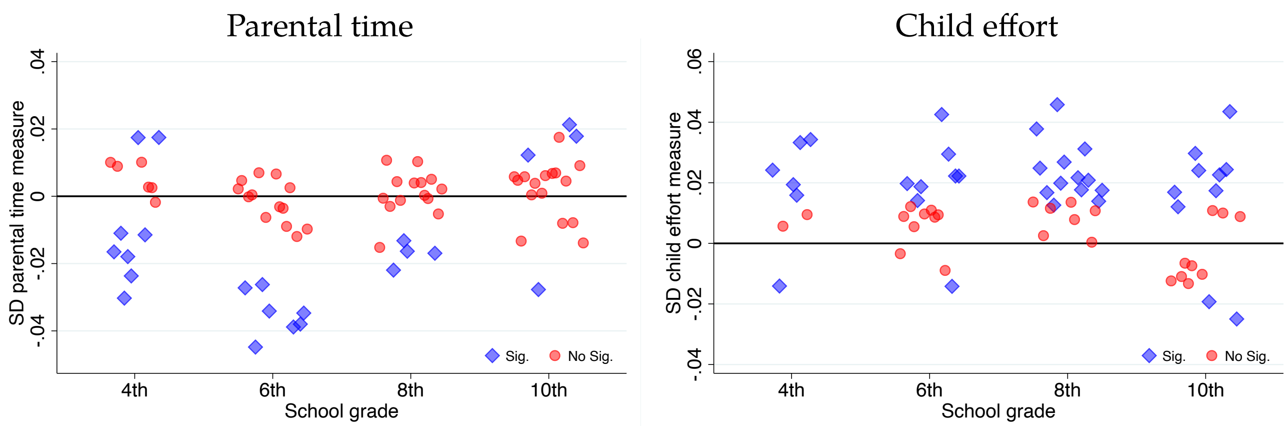
$$A_{ijt} = X'_{it}\beta + \mathrm{V}_{jt} + \varepsilon_{ijt}$$

where X_{it} is a vector of the student's characteristics including a second order polynomial in previous test scores, student's gender and age, parents' education and age, household income, indicator of mother answered questionnaire and indicators of missing controls. The V_{jt} are classroom fixed effects and ε_{ijt} is an error term. The estimates of V_j are the measures of classroom value-added.

[31] I regress each time investment measure on classroom value-added conditioning of a set of variables: second order polynomials of previous scores, household income, student's gender and age, parents' education and age, and school fixed effects. The sample is restricted to students in schools with at least two classrooms.

across questions, either within or between school grades. Nevertheless, the sign of the coefficient helps us understand the direction of the household response to the classroom effects. The left panel of the figure shows the coefficients for parental time. The coefficients are mostly negative for grades 4 to 8 while for grade 10 the coefficients are largely positive. This suggests that parents respond differently depending on the grade their child is attending. The right panel shows the same estimates but for child time investment. The coefficients are mainly positive at every school grade indicating that children reinforce classroom effects.

Figure 1.1:
Ordered categorical questions of time investment and classroom value-added



Note: The sample consists of students at schools with at least two classrooms. Each dot presents the coefficient of a regression of parental time or child effort (standardized) ordered categorical variable on classroom value-added. All specifications control for second order polynomials of previous scores, parents' education and age, student's age and gender, household income (second order polynomial), and school fixed effects.

The main drawback of this approach is the interpretation and comparability of the estimates across questions. The coefficients are informative about the direction of the response, but their magnitudes are difficult to interpret. This does not allow an analysis of the evolution of the responses across school grades or their magnitude. To deal with this problem it is necessary to formalize a measurement system. In the next section, I describe the measurement strategy that I follow.

1.4 Measurement of skills and time investments

In this section, I describe the measurement of the cognitive skills and the time investment of households. For skills I consider a linear (or log-linear) system of measures for the latent skills stocks. And for time investment, I build a non-linear response model for ordered categorical questions.

1.4.1 Skill measurement

The cognitive skills of children are not directly observable. Following Cunha and Heckman (2008), I assume test scores are arbitrary scaled measures of latent cognitive skills. Test scores are denoted by M_{itm}, measures (test scores in different subjects) are indexed by m and the measurement structure is:

$$M_{itm} = \mu_{tm} + \lambda_{tm} \log \theta_{it} + \varepsilon_{itm} \tag{1.4.1}$$

where θ_{it} is the skill level of child i at school grade t, μ_{tm} and λ_{tm} are the parameters of the measurement system and ε_{itm} is a zero mean error term. Assuming that the measures are linearly related to the natural logarithm of skills constrains the skill values to be positive numbers, which is required given the functional form assumed for the skill formation technology in equation (1.2.6). The parametric assumption of the technology and its implications for the measurement system are without loss of generality for the estimation and interpretation of household responses. However, these assumptions do affect the interpretation of the skill technology and the preference parameters.

1.4.2 Time investment measurement

Existing work has used linear and non-linear measurement systems for latent investments.[32] Typically, observed measures of investments are ordered categorical variables with two to five categories. Linear models are estimated under the assumption that these ordered categorical questions are continuous measures. Instead, non-linear models are built on the discrete structure of the questions at the expense of imposing functional form assumptions on the distribution of error terms. The objective of these measurement systems is to identify and estimate parameters characterizing the relationship between the latent investment and other variables—e.g., skills and classroom inputs. Whether the latent factor is a dependent or independent variable in the analysis has relevant implications regarding the assumptions of the measurement system.

I use a non-linear measurement system for time investments. The main reason is that the categorical ordered questions of time investments are not consistent across grades or calendar years. Assuming a linear model requires re-normalizations to identify the system, which places stringent assumptions on the system's parameters (Agostinelli and Wiswall, 2016). However, a non-linear system is not necessarily a sub-optimal approach, it depends on the trade-off between additional error due to the continuity assumption and misspecification.[33] A nice feature of the non-linear approach is that it estimates the time investment of each student measured in hours, as opposed to standard deviations, leading to a clearer interpretation of the parameters of interest.

The questions of time investment in the administrative data are denoted by Z_{its} and different questions are indexed by s. The total number of questions at grade t is labeled by S_t. These are ordered categorical variables—i.e., $Z_{its} \in \{1, 2, \ldots, K_s\}$ where K_s denotes

[32] Cunha and Heckman (2008), Cunha et al. (2010), and Agostinelli et al. (2020) use linear systems, while Fu and Mehta (2018), Agostinelli (2018), and Wang (2020) use non-linear measurement models.

[33] Appendix C presents a Monte Carlo simulation in which I evaluate the asymptotic properties of linear and non-linear models. In the linear case, the consequence of treating categorical variables as continuous results in *additional* measurement error. In the non-linear case, estimates are more efficient but this is not necessarily true under misspecification. Both alternatives produce consistent estimators of the parameters of interest, but neither choice is preferred *a priori* in terms of efficiency.

the number of categories of measure s.[34] I assume a multivariate ordered response model where the time investment h_{it} (measured in hours) is the latent variable. The structure of the response rule is:

$$Z_{its} = k \quad \text{if and only if} \quad \alpha_{stk} \leq \beta_{st} h_{it} + \epsilon_{its} < \alpha_{stk+1} \qquad \text{for } k = 1, 2, \ldots, K_s, \tag{1.4.2}$$

where $\alpha_{st1} = -\infty$ and $\alpha_{stK_s+1} = \infty$. I assume that the error terms ϵ_{its} have logistic distributions.[35] Note that the parameters in the measurement system are indexed by question and school grade. Typically, latent variables do not have a natural scale and location or known distribution. Researchers address this with a normalization and by assuming the distribution of the latent variable. However, parental and child time investment have a ratio scale, the only degree of freedom is the units—e.g., weekly or daily hours.[36] Even though these parameters are identified within grade, the properties of the scale of the latent variable allows comparisons across school grades.

1.5 Identification of measurement systems, skill technology, and policy functions

In this section, I present the identification analysis of the parameters of the measurement systems of cognitive skills and time investment, the skill formation technology, and the approximations of the time investment policy functions.

[34]See Table S.1.2 for examples of the structure of these variables.

[35]This is similar to models of the item response theory literature. In graded response models each measure has its own discrimination parameter β_{st} and thresholds $\{\alpha_{stk}\}_{k=1}^{K_s}$, which identify boundaries between the ordered outcomes (Samejima, 1969).

[36]Ratio scale implies that the ratio of two intervals is unit free and the measure has natural origin.

1.5.1 Measurement systems

I build a measurement system for the cognitive skills of children and the time investment of households. First, I describe the dynamic measurement system of cognitive skills where the measures are given by the standardized test scores in the administrative data and the WISC-V cognitive test survey. Second, I present the measurement system for time investment of households using the ordered categorical questions of time investment from the questionnaires filled out by parents and children in the administrative data.

Skill measurement

This section describes the identification of the skill measurement system. Driven by the structure of the data, I assume there are two kinds of skill measures: age-invariant and age-varying measures.[37] In addition, the joint distribution of the age-invariant skill measures and any other variable (including the age-varying measures) is not observed, whereas the joint distribution of age-varying measures and other variables is observed.

The skill measurement system is a dynamic linear (or log-linear) latent factor model. Existing work in the child development literature has developed two different strategies to identify the dynamic system: 1) Placing parametric assumptions on the skill technology and anchoring skills' scales to outcomes in adulthood (Cunha and Heckman, 2008; Cunha et al., 2010); or 2) using age-invariant measures of skills (Agostinelli and Wiswall, 2016). I do not follow these strategies because the administrative data does not have age-invariant measures or future outcomes of children, such as earnings.

With only age-varying measures available it is not possible to identify the skill technology's parameters across grades without re-normalization (Agostinelli and Wiswall,

[37] I follow the definition of age-invariant measures in Agostinelli and Wiswall (2016). This definition follows from a line of research in psychometrics that aims to measure children's cognitive development and track its evolution as they grow up. That is, it aims to measure the cognitive development regardless of the age of children. The term "age-varying measures" is not quite accurate. It is not that the measures themselves are age-varying, they are technically different across grades—e.g., math tests are designed differently at every grade. However, I follow this nomenclature to facilitate the comparison between the two kinds of measures.

2016). Nevertheless, the system within grades is identified by exploiting orthogonality conditions under usual assumptions (Cunha and Heckman, 2008). Since the household responses are estimated through variation within grades, re-normalization on the skill measurement system does not prevent the identification of the coefficients of the time investments functions—i.e., equation (1.2.9). However, re-normalization implies that the skill technology is not comparable across grades. This restricts the analysis of the dynamics of the skill formation process and it frustrates the identification of the preference parameters of the child development model. To overcome this problem, I use additional data from the WISC-V cognitive development test survey, which has age-invariant measures and is representative of the same population. This information allows me to track the evolution of the skill distribution across grades and to identify the dynamic measurement system.

Let M^A_{itm} be a age-invariant measure of skills, where i, t, and m index children, their age/grade, and measure, respectively. These are the measures available in the WISC-V cognitive development test survey. Their structure is as follows:

$$M^A_{itm} = \mu^A_m + \lambda^A_m \log\theta_{it} + \varepsilon_{itm}, \tag{1.5.1}$$

where θ_{it} is the skill of the child and ε_{itm} is a mean zero error term. Note that the parameters μ^A_m and λ^A_m are not indexed by t—i.e., the measures are age-invariant. I assume the usual independence assumption about the error terms to identify linear latent factor models:

Assumption 1:

- $\varepsilon_{itm} \perp \varepsilon_{itm'}$ for all t and all $m \neq m'$;
- $\varepsilon_{itm} \perp \log\theta_{it}$ for all m and all t.

Skills do not have a natural scale and location and consequently identification requires a normalization. At this point, it is important to remember that the measures are assumed to proxy for the logarithm of skills due to the parametric assumption of the technology. Thus, it follows that the normalization is over the scale of the latent factor $\log\theta_{it}$, rather than θ_{it}.

That is:

Normalization 1:

- $\mathrm{E}(\log\theta_{i0}) = 0$;
- $\mathrm{Var}(\log\theta_{i0}) = 1$.

The normalization is for the variance and mean of the log skill distribution at an arbitrary age $t = 0$. I set $t = 0$ to be the age of 8 years old when children attend second grade. Then, exploiting the orthogonality conditions in Assumption 1, the parameters are identified using expectation and covariance moments through the following equations:

$$\mu_m^A = \mathrm{E}(M_{i0m}^A); \qquad \lambda_m^A = \sqrt{\frac{\mathrm{Cov}(M_{i0m}^A, M_{i0m'}^A)\mathrm{Cov}(M_{i0m}^A, M_{i0m''}^A)}{\mathrm{Cov}(M_{i0m'}^A, M_{i0m''}^A)}}. \tag{1.5.2}$$

Once these parameters are identified, it is possible to identify the expected value and variance of the latent skill for children at each grade t:

$$\mathrm{E}(\log\theta_{it}) = \frac{\mathrm{E}(M_{itm}^A) - \mu_m^A}{\lambda_m^A}; \qquad \mathrm{Var}(\log\theta_{it}) = \frac{\mathrm{Cov}(M_{itm}^A, M_{itm'}^A)}{\lambda_m^A \lambda_{m'}^A}. \tag{1.5.3}$$

The evolution of these moments allows me to identify the parameters of the skill measurement system of the age-varying measures, M_{itm}. These measures have similar structure as the age-invariant measures but their parameters vary by the age of children:

$$M_{itm} = \mu_{mt} + \lambda_{mt} \log\theta_{it} + \varepsilon_{itm}, \tag{1.5.4}$$

where the error terms are assumed to have the same properties defined in Assumption 1. The only difference is that the parameters μ_{mt} and λ_{mt} are indexed by t. Since $\mathrm{Var}(\log\theta_{it})$ and $\mathrm{E}(\log\theta_{it})$ are identified, λ_{mt} and μ_{mt} can be identified using the orthogonality conditions:

$$\begin{aligned} \lambda_{mt} &= \sqrt{\frac{1}{\mathrm{Var}(\log\theta_{it})} \times \frac{\mathrm{Cov}(M_{itm}, M_{itm'})\mathrm{Cov}(M_{itm}, M_{itm''})}{\mathrm{Cov}(M_{itm'}, M_{itm''})}} \\ \mu_{mt} &= \mathrm{E}(M_{itm}) - \lambda_{tm}\mathrm{E}(\log\theta_{it}). \end{aligned} \tag{1.5.5}$$

Thus, the latent skills are identified up to some measurement error through the following equation:

$$\tilde{M}_{itm} = \frac{M_{itm} - \mu_{mt}}{\lambda_{mt}} = \log \theta_{it} + \frac{\varepsilon_{itm}}{\lambda_{mt}}. \tag{1.5.6}$$

It is important to note that the operation in equation (1.5.6) is an affine transformation of the test scores. Thus, the original scale of the test scores is preserved.[38]

Time investments measurement

This section explains the identification regarding the parameters of the response model of the time investments questions and the identification of each student's time investment. The identification of the model parameters follows from the expected outcomes of the ordered categorical questions of time investment and the distribution of time investment at each school grade t, denoted by $g_t(\cdot)$.[39] I follow San Martín et al. (2013) for the identification of the response model's parameters. I assume there are at least three questions available within a grade, $S_t \geq 3$, and that:

Assumption 2:

- $\epsilon_{its} \perp h_{it}$ for all s and all t;
- $\epsilon_{its} \perp \epsilon_{its'}$ for all t and all $s \neq s'$.

To simplify notation and without loss of generality, I drop the school grade index t and take the case that $K_s = 2$ for all s and that Z_{is} is equal to 0 or 1. Let $\Pr(Z_{is} = 1 \mid h_i; \alpha_s, \beta_s)$ be the probability of $Z_{is} = 1$, conditional on the time investment. Note that for $K_s = 2$ for all s the response model has two parameters for each measure s: α_s and β_s. The expected value of observing outcome 1 for measure s is:

$$p_s \equiv \Pr(Z_{is} = 1) = \int \Pr(Z_{is} = 1 \mid h_i; \alpha_s, \beta_s) g(h_i) dh_i, \tag{1.5.7}$$

[38] In value-added models of academic achievement it is usual to use a standardized version of test scores. These procedures are affine transformation as well and preserve the original scale of the test scores.

[39] In item response theory, where the latent variable is student *ability*, it is assumed that ability is distributed as a standard normal. The advantage of the current framework is that parental and child time investment have ratio scales and their marginal distributions can be estimated from the time use survey.

Given $g(\cdot)$ and p_s, for any β_s the function $\Pr(Z_{is} = 1)$ is strictly decreasing in α_s, so the inverse with respect to α_s exists and is unique. So we can define $\alpha_s = \alpha(\beta_s, p_s)$. The joint probability of measures s and s' being both equal to one is:

$$
\begin{aligned}
p_{s,s'} &= \Pr(Z_{is} = 1, Z_{is'} = 1) \\
&= \int \Pr(Z_{is} = 1 \mid h_i; \alpha(\beta_s, p_s), \beta_s) \times \Pr(Z_{is'} = 1 \mid h_i; \alpha(\beta_{s'}, p_{s'}), \beta_{s'}) g(h_i) dh_i \\
&\equiv q(\beta_s, p_s, \beta_{s'}, p_{s'}).
\end{aligned}
\tag{1.5.8}
$$

Then, since $q(\cdot)$ is strictly increasing in β_s and $\beta_{s'}$ (see Appendix B for the proof), we can define $\beta_{s'} = \overline{q}(\beta_s, p_s, p_{s,s'}, p_{s'})$. If there are at least three time investment questions—i.e., $S \geq 3$—we have for $s = 1, 2, 3$:

$$
\begin{aligned}
p_{1,2} &= q(\beta_1, p_1, \beta_2, p_2) \\
p_{1,3} &= q(\beta_1, p_1, \beta_3, p_3) \\
p_{2,3} &= q(\beta_2, p_2, \beta_3, p_3).
\end{aligned}
\tag{1.5.9}
$$

It follows that:

$$
p_{2,3} = q(\overline{q}(\beta_1, p_1, p_{1,2}, p_2), p_2, \overline{q}(\beta_1, p_1, p_{1,3}, p_3), p_3) \equiv r(\beta_1, p_1, p_2, p_3, p_{1,2}, p_{1,3}). \tag{1.5.10}
$$

Lastly, since $r(\cdot)$ is strictly decreasing in β_1 (Appendix B presents the proof), it follows that the inverse of $r(\cdot)$ on β_1 exists and so β_1 is identified. Similarly, β_s and α_s are identified for all s. Note that without knowledge of $g_t(\cdot)$ identification of this parameters is not possible. In the empirical application I estimate this distribution from additional data from the Chilean Time Use Survey.

Besides identification of the parameters of the response model, we need to identify each student's time investment. This is a difficult and complex task. In a similar framework, the item response theory has developed several strategies to identify students' latent *ability* from their answers to a series of questions. Both time investment and ability are continuous variables and the objective is to identify a point in their support from a set variables with discrete and finite support. Any attempt of point identification requires the number of questions to go to infinity. However, even under this condition, $S \to \infty$, identification is not achieved under several natural strategies—i.e., using moments from the individual likelihood or posterior distribution. For example, Lord (1980) suggested identifying the latent value of each individual up to some error with the mode of the individual likelihood or the mode of the conditional posterior distribution. The difference between these moments and the individual's latent value are of order S^{-1}—i.e., $O(S^{-1})$.

Mislevy (1991) develops a different approach. He focuses on the identification of parameters that relate the latent variable and other variables—e.g., the effect of time investment on cognitive skills or the effect of classroom inputs on time investments.[40] His work has been widely used and extended (Mislevy, 1993; Junker et al., 2012; Schofield, 2014; Schofield et al., 2015). Mislevy proposed using plausible values but from the posterior distribution conditional on the ordered categorical questions and additional variables. The additional variables are those whose relationship to the latent variable is of interest.[41] Using draws from this distribution secures identification of the parameters that relate the latent variable and other variables. However, this strategy requires identifying the latent variable distribution conditional on the questions and additional variables. In my framework, this strategy is not possible, since the distribution of the time investment is identified from the Chilean Time Use Survey and, for example, it does not have test scores or the students' classroom assignment to identify the conditional distribution.

I follow the identification strategy proposed by Warm (1989). In particular, he suggested identifying the latent value of each student as the mode of the individual likelihood times

[40]In the same vein, Williams (2019) uses a semiparametric approach to identify the relationship of two variables while "controlling" for the latent variable.

[41]Plausible values are defined as random draws from the latent variable's posterior distribution. The plausible values (Rubin, 1987) differences from the true latent value are $O(S^{-1})$ and the parameters of the relationship between the latent and other variables are not identified in general.

a weighting function as:

$$\tilde{h}_{it} \equiv \underset{h_{it}}{\operatorname{argmax}}\; w(h_{it}) \times f(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta) \tag{1.5.11}$$

where $w(h_{it})$ is the weighting function. By definition, the derivative of the function of the right side of equation (1.5.11) is equal to zero at the mode:

$$\frac{\partial \log f(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta)}{\partial h_{it}} + \frac{\partial \log w(h_{it})}{\partial h_{it}} = 0. \tag{1.5.12}$$

The first term of the equation (1.5.12) is the derivative of the individual log likelihood. Note that if the weighting function is equal to one, the solution of this equation is the mode of the individual likelihood while if the weighting function is the prior distribution—i.e., the time investment distribution—the result is the mode of the posterior conditional distribution. Warm suggested as the weighting function a function such that its derivative with respect to the time investment is a function $\mathrm{B}(h_{it})$, defined as the difference between the mode of the individual likelihood and the time investment, times the Fisher information function $\mathrm{I}(h_{it})$:[42][43]

$$\frac{\partial \log f(\mathbf{z}_{it} \mid \tilde{h}_{it})}{\partial h_{it}} + \mathrm{B}(\tilde{h}_{it}) I(\tilde{h}_{it}) = 0. \tag{1.5.13}$$

Then, Warm proves that $\tilde{h}_{it} \to h_{it}$ as $S \to \infty$—i.e., the asymptotic difference between $\tilde{h}_{it}$ and h_{it} is $o(S^{-1})$. However, since (α, β) are not observed, the identification of h_{it} is up to some error. I define this error as ζ_{itl} and the index l indicates the set of variables $\mathbf{Z}_{it}$ included in the identification of h_{it}. These error terms are i.i.d. with mean zero and finite variance.

[42] Warm (1989) considered dichotomous ability measures and Samejima (1993) provided an extension of the second term in Warm's equation (1.5.12) for polytomous measures, such as these in the current setting.

[43] Even though $w(h_{it})$ does not have closed form, both $\mathrm{B}(h_{it})$ and $\mathrm{I}(h_{it})$ do have closed form; I provide the analytic functions in Appendix B.

1.5.2 Skill formation technology

The identification of the skill technology closely follows Agostinelli et al. (2020). I plug in the noisy measures of skills and time investment defined by equations (1.5.6) and (1.5.13) in equation (1.2.6) and rearrange:

$$\tilde{M}_{it+1m} = \log A_t + \gamma_{1t}\tilde{M}_{itm} + \gamma_{2t}\tilde{h}_{itl} + \gamma_{3t}\log C_{it} + \gamma_{4t}\tilde{h}_{itl} \times \log C_{it} + \mathbf{z}'_{it}\gamma_{5t} + \tilde{\nu}_{it}, \quad (1.5.14)$$

where the term $\tilde{\nu}_{it}$ includes the structural shock and additional measurement error; in particular:

$$\tilde{\nu}_{it} \equiv \nu_{it} - \frac{\gamma_{1t}\varepsilon_{itm}}{\lambda_{mt}} - \gamma_{2t}\zeta_{itl} - \gamma_{4t}\zeta_{itl} \times \log C_{it} + \frac{\varepsilon_{it+1m}}{\lambda_{mt+1}}. \quad (1.5.15)$$

There are two threats to identification of the parameters in equation (1.5.14). One is the presence of measurement error, and second is the potential correlation between the variables on the right-hand side with the structural shock ν_{it}. I identify the classroom effects as the natural logarithm of classroom inputs. As in the case of the skills, this follows from the functional form assumption of the technology and it is without loss of generality regarding the identification of the household responses. However, it does matter for the interpretation of the parameters of the skill technology and preferences. Moreover, since classroom effects do not have a natural scale or location, I perform the following normalization:

Normalization 2:

- $\mathrm{E}[\log C_{it}] = 0$ and;
- $\mathrm{Var}[\log C_{it}] = 1$ for every t.

The identification of the parameters under the presence of measurement error requires exclusion restrictions from additional noisy measures of time investments and skills. Besides Normalization 1 and Assumption 1 and 2, I require the following assumptions on the error terms for exclusion conditions to be valid:

Assumption 3: For $d_{it} \in \{\log C_{it}, \log \theta_{it}, h_{it}, \mathbf{z}_{it}\}$ and $\omega_{itj} \in \{\varepsilon_{itm}, \zeta_{its}\}$ where $j = m, s$.

- $\omega_{itj} \perp \omega_{it'j}$ for all j and $t \neq t'$;

- $\omega_{itj} \perp \omega_{it'j'}$ for every $t \neq t'$ and all j and j';
- $\omega_{it} \perp \omega'_{it'}$ for all $\omega \neq \omega'$ and all t and t';
- $\omega_{itj} \perp d_{it}$ for all j and all t;
- $\omega_{itj} \perp \nu_{it'}$ for all j and all t and t'.

The last assumption is mean-independence of the structural shock:

Assumption 4: mean-independence:

$$\mathrm{E}[\nu_{it} \mid \log C_{it}, \log \theta_{it}, h_{it}, \mathbf{z}_{it}] = 0. \tag{1.5.16}$$

Then, under Normalization 1 and 2 and Assumptions 1 to 4 the parameters of equation (1.5.14) are identified. The proof is the usual one in the framework of fixed effects models with instrumental variables. Assumption 4 is fundamentally not testable. However, I follow Chetty et al. (2014a) and perform an indirect test using omitted observable variables, such as household income. Moreover, I evaluate out-of-sample prediction performance of the skill technology as in Agostinelli et al. (2020). I describe these tests and the results in detail in the estimates section.

Additionally, under Assumptions 1 to 4 and Normalizations 1 and 2, the classroom effects are identified up to some error. Since the classroom effect measures have error, we require additional measures to exploit orthogonality conditions between error terms to identify the household responses. Thus, using disjoint random groups of students at each classroom, I identify classroom effects with different error terms. The identified classroom effect is defined as:

$$\log \tilde{C}_{itc} = \log C_{it} + \chi_{itc}, \tag{1.5.17}$$

where χ_{itc} is the error associated with a particular sample of students, indexed by c.

1.5.3 Time investment functions

This section describes the identification of the time investment functions. I plug in the noisy counterparts of time investment, classroom effects, and skills defined by equations (1.5.6), (1.5.13), and (1.5.17) into equation (1.2.9) and rearrange:

$$\begin{aligned} \tilde{h}_{itl} &= \delta_{0,t} + \delta_{C,t} \log \tilde{C}_{itc} + \delta_{\theta,t} \tilde{M}_{itm} + \mathbf{\Gamma}'_{it} \delta_{\mathbf{\Gamma},t} + \pi_i + \tilde{\eta}_{it}, \\ & \qquad \qquad \qquad \text{where} \quad \tilde{\eta}_{it} \equiv \eta_{it} - \delta_{C,t} \chi_{itc} - \delta_{\theta,t} \frac{\varepsilon_{itm}}{\lambda_{tm}} + \zeta_{itl}. \end{aligned} \tag{1.5.18}$$

As in the case of the skill technology, there are two threats to identification of the parameters in equation (1.5.18): measurement error and selection on unobservables. The former is addressed by exploiting exclusion conditions using additional skill and classroom effects measures.[44] These exclusion restrictions hold if *additionally* to Normalization 1 and 2 and Assumptions 1 to 4, the following assumption is met:

Assumption 5: For $d_{it} \in \{\log C_{it}, \log \theta_{it}, h_{it}, \mathbf{z}_{it}\}$ and $\omega_{itj} \in \{\varepsilon_{itm}, \zeta_{itl}\}$ where $j = m, l$.

- $\chi_{itc} \perp \chi_{it'c}$ for all j and $t \neq t'$;
- $\chi_{itc} \perp \omega_{it'j}$ for every t and t' and all c and j;
- $\chi_{itc} \perp d_{it}$ for all c and all t;
- $\chi_{itc} \perp \eta_{it'}$ for all c and all t and t'.

Finally, the last assumption required is mean-independence—i.e., η_{it} is mean zero conditional on current skills, classroom inputs, observable characteristics, and all individual time-invariant unobservables. Formally,

Assumption 6: mean-independence:

$$\mathrm{E}[\eta_{it} \mid \log C_{it}, \log \theta_{it}, \mathbf{\Gamma}_{it}, \pi_i, t] = 0. \tag{1.5.19}$$

Under Normalization 1 and 2 and Assumptions 1 to 5 the parameters in equation (1.2.9) are identified. The proof is the usual one in the framework of fixed effects models with

[44]The additional skill measures are provided by test scores of different subjects and the additional measures of classroom effects are generated using disjoint random groups of students to identify the classroom effects.

instrumental variables. Note that if we only care about identification of $\delta_{C,t}$ we only require conditional mean independence, which is a weaker assumption.

1.6 Estimation of measurement systems, skill technology, and policy functions

In this section I describe the estimation procedure, which consists of three blocks. The first block is the estimation of the measurement systems of skills and household time investments. The output is noisy measures of each student's time inputs and skills. The second block consists of the estimation of the skill formation technology that, as a by-product, provides estimates of classroom and teacher effects. Lastly, I estimate the time investment functions that provide the household response to school inputs.

1.6.1 Measurement system

In the current section, I present the estimation strategy of the measurement system of cognitive skills and time investment of parents and children.

Skill measurement

The skill measurement system is estimated by replacing the moments in equations (1.5.2), (1.5.3) and (1.5.5) with their sample analogs. Note that using different combinations of measures provides several estimates of each parameter, the final estimate corresponds to the average across the estimates of different combination of measures. Using these estimates, I implement the affine transformation over the test scores given in equation (1.5.6).

Time investment measurement

The time investments questions in the administrative data—i.e., both parental and child time investment—are ordered categorical questions. Under the assumptions made over the response rule and error terms, I build the sample likelihood and estimate the parameters of the system using a maximum likelihood estimator. The probability of observing outcome k or higher for measure s, conditional on the individual time investment, is given by:

$$\Pr(Z_{its} \geq k \mid h_{it}) = \frac{\exp(\beta_{st} h_{it} - \alpha_{stk})}{1 + \exp(\beta_{st} h_{it} - \alpha_{stk})}. \tag{1.6.1}$$

The conditional probability of observing outcome k can be expressed as the difference of two of the previous probabilities,

$$\Pr(Z_{its} = k \mid h_{it}) = \Pr(Z_{its} \geq k \mid h_{it}) - \Pr(Z_{its} \geq k + 1 \mid h_{it}), \tag{1.6.2}$$

where $\Pr(Z_{its} \geq 0 \mid h_{it}) = 1$ and $\Pr(Z_{its} > K_s \mid h_{it}) = 0$. Let z_{its} be a possible response of Z_{its}. Given the independence assumption across the error terms of different questions, the density function for a student, conditional on h_{it}, is:

$$f(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta) = \prod_{s} \Pr(Z_{its} = z_{its} \mid h_{it}), \tag{1.6.3}$$

where $\mathbf{z}_{it} = (z_{it1}, \ldots, z_{itS_t})$. The likelihood for a single student is computed by integrating out the latent variable—i.e., time investment—from the joint density,

$$L_{it}(\alpha, \beta \mid \mathbf{z}_{it}) = \int f(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta) g_t(h_{it}) dh_{it}, \tag{1.6.4}$$

where $g_t(\cdot)$ is the density function of the time investment. I estimate density function $g_t(\cdot)$ using the Chilean Time Use Survey of 2015. There are two assumption required to perform this step: 1) the population in both databases are the same—i.e., the administrative data and the time use survey; and 2) the time input distribution at each school grade is stable during the time frame of the analysis. The first assumption is fairly weak, since the administrative data is virtually a census of children at each school grade and the survey is nationally representative. The main concern comes from missing data. The missing information rate is around 15 percent in the administrative data and less than 5 percent in the time use survey.[45] This is not a problem as long as the sample selection is independent or similar in both samples. The second assumption is made due to data availability since the time use survey was only collected in 2015. However, the entire period covered in the analysis is from 2011 to 2018, and there was not a significant event that might have dramatically affected the time investment distribution during this period.

The log likelihood of the sample is simply the sum of the log likelihoods of the N_t students at the school grade t:

$$\log L_t(\alpha, \beta \mid \mathbf{z}_{it}) = \sum_{i=1}^{N_t} \log L_{it}(\alpha, \beta \mid \mathbf{z}_{it}), \tag{1.6.5}$$

and the maximum likelihood estimator is defined as:

$$(\hat{\alpha}, \hat{\beta}) \equiv \underset{(\alpha,\beta)}{\text{argmax}} \log L_t(\alpha, \beta \mid \mathbf{z}_{it}). \tag{1.6.6}$$

As the estimation procedure, I use the Marginal Maximum Likelihood/EM approach of Bock and Aitkin (1981). Once these parameters are estimated, the response model can be use to estimate each household's time inputs. The item response theory literature develops several strategies in the context of students' *ability*. The most common estimators are Bayesian or frequentist. The former estimates each individual latent value using statistics of the posterior conditional distribution, such as the expected value or mode. The frequentist approach estimates the latent value by maximizing each individual's likelihood. Both strategies result in biased estimators where the bias is a function of the level of the latent factor. While the Bayesian approach compresses the estimated latent variable's distribution towards the prior distribution's expected value, the maximum likelihood approach tends

[45] The administrative data on time investments is collected via questionnaires to parents and students, which are not mandatory.

to spread the estimates' distribution relative to the distribution of the true latent variable across individuals.

Existing work that uses non-linear measurement models for time investments, such as Agostinelli (2018) and Wang (2020), relies on Bayesian strategies. Agostinelli (2018) uses indicator variables of parents spending time with children in specific activities. He assumes a uniform distribution as the prior distribution of the probability of a positive answer and estimates the posterior distribution using the conjugate multinomial using the answers to the dichotomous questions. Then, he draws realizations of the probabilities from the estimated posterior. The time investment of each child is estimated by applying the inverse of the parental time unconditional distribution (estimated from a time use survey) to these drawn probabilities.[46] In contrast, Wang (2020) uses a multivariate ordered response model similar to the one presented in sections 1.5. She uses the model to estimates the posterior distribution (augmented with covariates) of household investments and takes draws from these distributions.

Both authors generate bias estimates of each individual's investments. This bias depends on the level of the investment and is $O(S^{-1})$. Their approaches are Bayesian and in consequence put some weight on the prior distribution. Students with parental time in the extremes of the distribution have a larger bias than those closer to the prior's mean. Wang's approach, however, ameliorates this issue by including covariates in her posterior distribution, resulting in estimates that are biased towards the expected value of the conditional distribution instead of the unconditional one. Wang is not interested in estimating the time allocation of individual parents, but rather moments of the conditional distribution of parental time. Thus, her estimates are not bias due to the use of this Bayesian methodology.

Fu and Mehta (2018) use an ordered response framework for parental effort in a general equilibrium model of school tracking. They build the likelihood of the data using the distribution of ordered categorical questions conditional on the true value of parental

[46] In item response theory this kind of estimators of individuals' latent variables are usually called plausible values (Rubin, 1987).

effort. They estimate the model's parameters without estimating each student's parental effort. I do not follow this strategy because the estimation of the skill technology and time investment functions require estimating a large number of classroom effects and students fixed components, respectively. It would imply estimating thousands of parameters or using a random effects approach which requires strong assumptions regarding the distribution of classroom effects and the student heterogeneity distribution.

Instead, I use a strategy proposed in Warm (1989). In the context of student ability, Warm (1989) developed an estimator that aims to correct for the bias in estimates of each individual's latent value, called the weighted maximum likelihood estimator. The estimation of each student's time investment uses as input the estimates of the model's parameters. In particular, replacing the maximum likelihood estimates, $\hat{\alpha}, \hat{\beta}$, in equation (1.5.13) and solving for $\tilde{h}_{itl}$. That is,

$$\frac{\partial \log f(\mathbf{z}_{it} \mid \tilde{h}_{itl}; \hat{\alpha}, \hat{\beta})}{\partial \tilde{h}_{itl}} + \mathrm{B}(\tilde{h}_{itl}; \hat{\alpha}, \hat{\beta}) I(\tilde{h}_{itl}; \hat{\alpha}, \hat{\beta}) = 0. \tag{1.6.7}$$

Note that $\tilde{h}_{itl}$ is estimated using the estimates $(\hat{\alpha}, \hat{\beta})$ instead of the true parameters. Then, the estimates $\tilde{h}_{itl}$ have additional estimation errors, denoted as ζ_{it}. I estimate the response model several times using disjoint sets of the categorical outcome questions Z_{its}. Each of these estimates provide time investment measures that have different estimation error ζ_{itl}, where each disjoint set of questions is indexed by l. It follows that $\zeta_{itl} \perp h_{it}$ for all t and l and $\zeta_{itl} \perp \zeta_{itl'}$ for all $l \neq l'$ and all t.

1.6.2 Skill formation technology

In this section, I describe the methodology I employ to estimate the technology of skill formation. In particular, I use an estimation strategy developed by Agostinelli et al. (2020). They provide an extension of the algorithm in Arcidiacono et al. (2012) which allows them to estimate classroom effects with interaction terms with observable inputs and exploit instrumental variables to correct for bias due to measurement error in inputs. The methodology estimates classroom effects as the systematic variation in the skills of students assigned to the same classroom, in a similar fashion as the education production function literature.

I use the algorithm to estimate the technology of each school grade separately. The algorithm is as follows: It starts by taking an initial guess of the parameters of the skill technology $\gamma_t^0 \equiv (A_t^0, \gamma_{1t}^0, \ldots, \gamma_{4t}^0, \gamma_{5t}^0)$. Each iteration $n \in \{0, 1, \ldots\}$ of the algorithm consists of computing the following steps:

Step 1: Taking as given the current parameter guess n, γ_t^n, compute the classroom effect as the average within-classroom residual in skills at grade t:

$$\log C_{it}^n = \frac{\sum_{i' \in c(i)} \left[M_{i't+1m} - \log A_t^n - \gamma_{1t}^n M_{i'tm} - \gamma_{2t}^n \tilde{h}_{i't} - \mathbf{z}'_{i't} \gamma_{5t}^n \right]}{\sum_{i' \in c(i)} \left[\gamma_{3t}^n + \gamma_{4t}^n \tilde{h}_{i't} \right]} \tag{1.6.8}$$

where $c(i)$ is the set of children in the classroom that child i attends at grade t (including child i).

Step 2: Taking as given the distribution of classroom effects $\log C_{it}^n$ from **Step 1**, estimate the skill technology using the noisy measures of skills and time inputs:

$$\tilde{M}_{it+1m} = \log A_t^{n+1} + \gamma_{1t}^{n+1} \tilde{M}_{itm} + \gamma_{2t}^{n+1} \tilde{h}_{itl} + \gamma_{3t}^{n+1} \log C_{it}^n + \gamma_{4t}^{n+1} \tilde{h}_{itl} \times \log C_{it} + \mathbf{z}'_{it} \gamma_{5t}^{n+1} + \tilde{\nu}_{it} \tag{1.6.9}$$

I estimate these parameters with 2SLS estimator using additional measures of skills and time inputs in the first stage.[47] This produces the parameters for a new iteration $n+1$. The iteration procedure stops when all of the parameters converge—i.e., $||\gamma_t^{n+1} - \gamma_t^n||^\infty \approx 0$. Otherwise the algorithm returns to **Step 1** with the updated set of parameters γ_t^{n+1}. Once convergence is achieved, it provides the classroom effects distribution. I run the estimation

[47]OLS estimation of the **Step 2** equation produces inconsistent estimates of the remaining parameters due to measurement error bias.

by (disjoint) random samples of 50 percent of the students in each classroom to estimate classroom effects with different estimation error to exploit exclusion restrictions in the estimation of the time investment functions.[48]

I used a modified version of this algorithm to estimate the effects of different components of the overall classroom effect, such as the teacher effect $\log T_{it}$ and the effect of other observable classroom characteristics r_{it}. In my sample, I do not observe a set of teachers in every school teaching at some point at a different school. Thus, it is not possible to identify teacher and school effects separately while also generating a global ranking of teacher effects as in Mansfield (2015) and Chetty et al. (2014a,b). A popular alternative is to estimate within-school teacher effects. However, I estimate teacher effects as a compound of the effects of the school and teacher to leverage the variation of students across schools; which is ultimately the effect that should matter to the households. As for classroom effects, identification requires normalization of the teacher effects—i.e., teacher effects are set to be on average (sum up to) zero as in the normalization of the classroom effects.

1.6.3 Time investment functions

This section presents the estimation methodology of the time investment functions. Plug the observed noisy measures of time investment, skill and classroom effects defined in (1.5.13), equations (1.5.6) and (1.5.17) into equation (1.2.9) and rearrange:

$$\tilde{h}_{itl} = \delta_{0,t} + \delta_{C,t} \log \tilde{C}_{itc} + \delta_{\theta,t} \tilde{M}_{itm} + \mathbf{\Gamma}'_{it} \delta_{\mathbf{\Gamma},t} + \pi_i + \tilde{\eta}_{it} \tag{1.6.10}$$

[48]Note that there is an implicit assumption that ν_{it} is not observed by parents and children when they are making their investment choices. This can be relaxed by allowing correlation between ν_{it} and η_{it} and using a control function approach as in Attanasio et al. (2020b). Under this framework, the estimation of the skill technology and time investment functions should is done in a single step—i.e., including the estimation of the time investment function in Agostinelli et al. (2020)'s estimator. The algorithm is initialized by assuming a guess of the parameters and of the distribution of η_{it} to include a control function in the estimation of the skill technology. Then, at each iteration, after estimating the skill technology and classroom effects $\log C_{it}$ it is possible to estimate the time investment functions. The residuals provide estimates of η_{it} that can be used in the subsequent iterations until convergence is achieved.

There are three concerns that could bias the estimates of the parameters of equation (1.6.10): (i) non-classical measurement error in time investment; (ii) classical measurement error in skills or classroom effects; and (iii) unobserved factors that systematically influence time inputs, skills and classroom inputs. The potential bias related to point (i) is addressed by implementing the Warm (1989) weighted maximum likelihood estimator. Either way, it should be noted that the estimated time investment of each student has measurement error. However, this error is independent, which implies a precision cost but not bias in the estimates.

Point (ii) considers the situation where, even if h_{it} is observed and the structural shock η_{it} is not correlated with the observed inputs, the OLS estimates of $\delta_{C,t}$ and $\delta_{\theta,t}$ could be biased due to measurement error. I use the additional measures of each of these variables to exploit exclusion restrictions to estimate the parameters of equation (1.6.10). The additional skill measures are generated using test scores in different subjects and the additional classroom effects are estimated using disjoint samples of students at each classroom.[49] Finally, point (iii) suggests that the unconditional correlation between classroom (teacher) effects and time inputs could be the result of other unobserved factors. To deal with this concern, I use a student fixed effect approach leveraging the panel structure of the data and include time varying covariates, such as household income.[50] Thus, I estimate equation (1.6.10) using a 2SLS estimator with student fixed effects.

Since the coefficients are grade-specific, I rewrite equation (1.6.10) with a vector of variables in fourth grade and the variables in the upper grades interacted with a school grade indicator. Define the vectors $\mathbf{X}_{it} = (\log \tilde{C}_{itc}, \tilde{M}_{itm}, \mathbf{\Gamma}'_{it})'$ and $\delta_t = (\delta_{C,t}, \delta_{\theta,t}, \delta'_{\mathbf{\Gamma},t})'$. The

[49] As mentioned before, the instruments for the classroom effects are generated by estimating the technology with half the students randomly selected in each classroom. This strategy follows the spirit of leave-one-out estimators. There are two classroom effects estimates for each child; in only one was the student included in the estimation. In order for the estimate to be a valid instrument, it should be the estimate in which the student was not included. That is, the instrument is the classroom effect of each student estimated without her/him.

[50] Controlling for household income could be problematic. The parents' earnings might be affected by the parental time choice. However, the results are not quantitatively or qualitatively affected by the inclusion of this variable. The underlying assumption is that changes in parental time correspond to changes in leisure time and not in working hours.

vector of variables to include in the 2SLS estimator is:

$$b_{it} = (1, \mathbf{X}_{it}, \{1(t = t'), 1(t = t') \times \mathbf{X}_{it}\}_{t'>4})', \tag{1.6.11}$$

with $t' = 4, 6, 8, 10$ and $1(t = t')$ as an indicator function equal to 1 if $t = t'$ and zero otherwise. Similarly, I define the vector of associated coefficients:

$$\delta = (\delta_{0,4}, \delta_4, \{\tilde{\delta}_{0,t'}, \tilde{\delta}'_{t'}\}_{t'>4})', \tag{1.6.12}$$

where $\tilde{\delta}_{0,t} = \delta_{0,t} - \delta_{0,4}$ and $\tilde{\delta}_t = \delta_t - \delta_4$. Next, define the *demean* transformation for a variable a_{it} as:

$$\ddot{a}_{it} = a_{it} - \overline{a}_i + \overline{a}, \tag{1.6.13}$$

where $\overline{a}_i$ is the average value across all grades in which student i is observed and $\overline{a}$ is the grand average. The first step of the 2SLS estimator consists of regressing the noisy measures in $\ddot{b}_{it}$ on their demean equivalent measures and $\ddot{\mathbf{\Gamma}}_{it}$. Then, the estimated coefficients are use to generate predicted values of those noisy measures. Lastly, the second step consists of estimating a regression of $\ddot{h}_{itl}$ on the predicted measures and $\ddot{\mathbf{\Gamma}}_{it}$ using OLS estimator.

1.7 Estimates of measurement systems, skill technology, and policy functions

In this section, I describe the estimates of the measurement systems, the skill technology and time investment functions. Since the estimation methodology involves several steps, I estimate standard errors and 95% confidence intervals using the bootstrap with 1000 replications. The data has three dimensions: students, school grade and classrooms. The bootstrap sampling requires tracking students' entire history across grades and their classmates' as well. Moreover, students in connected schools might share similar shocks.

Thus, I cluster the bootstrap samples at the school network level. I define school networks as schools from which at least 20 students move between schools during the period of analysis. Note that at every bootstrap iteration, I estimate every step—i.e., measurement systems, skill formation technology, and time investments functions.

1.7.1 Measurement systems

Figure S.1.2 shows the estimates of the expected value of the log skills and its variance for ages between 8 and 16 years old. These estimates correspond to the empirical analogs of equation (1.5.3) estimated using the WISC-V cognitive test survey. Children's cognitive development improves as they grow up and its variance is larger for all ages relative to 8 and 16 years old where the variance is similar. Table S.1.3 presents the estimates of the parameters of the measurement system for the age-invariant measures and the signal share of each measure—i.e., the fraction of the measure's variance that is not attributed to the error.[51] Table S.1.4 shows the estimates and signal share for the measurement system estimated using the test scores from the administrative data. The signal share of these measures is similar to previous work using other databases.

1.7.2 Skill formation technology

Tables S.1.5 and S.1.6 present the estimates of the skill formation technology. The former shows the specification with the composite classroom input, while the later reports the within classroom inputs specification—i.e., teacher effects and observable classroom characteristics. These specifications are more general than those described in the previous sections. First, they include two sources of household time investment—i.e., parental and child time investment relabeled as h_{it} and e_{it}, respectively. Second, they include interactions of time inputs with current skills, between time inputs and between classroom inputs and current skills. This parameterization provides flexible heterogeneous effects of each input.

[51]The definition of the signal share is $1 - \text{Var}(\varepsilon_{itm})/\text{Var}(M_{itm})$. See equation (1.4.1) for more details.

Consider parental time's marginal effect for a particular student:[52]

$$\frac{\partial \log \theta_{it+1}}{\partial h_{it}} = \gamma_{3t} + 2\gamma_{4t} h_{it} + \gamma_{9t} \log C_{it} + \gamma_{11t} e_{it} + \gamma_{12t} \log \theta_{it}. \tag{1.7.14}$$

The parameters' signs in equation (1.7.14) define how inputs relate in production with parental time. For example, if $\gamma_{9t} > 0$, the effect of parental time is higher for children with better classroom inputs.[53] In the case of parental time and classroom inputs these parameters are positive for all grades except for sixth grade, where it is negative but not statistically significant. This means that at least for fourth, eight, and tenth grade there is complementarity in production between these inputs.

From these tables it is difficult to get a sense of the magnitude of each input's effect. Figure 1.2 plots the sample average of predicted values of equation (1.7.14) for each input. The top panel presents the average effect of parental (left) and child (right) time investment while the bottom panel shows the average effect of classrooms (left) and current skills (right). The grey areas represent school network-clustered bootstrap 95% confidence intervals. The average sample effect of one weekly hour of parental time is around 0.02 SD (of the log skills in second grade) at fourth grade and the effect decreases to 0.01 SD at tenth grade. The sample average effect of child time investment is higher at every grade but shows a decreasing pattern as well. The sample average marginal effects at fourth and tenth grade are 0.08 SD and 0.03 SD, respectively. Classroom inputs and current skills in

[52] The notation of this specification is as follows:

$$\begin{aligned}\log \theta_{it+1} \quad = \quad & \log A_t + \gamma_{1t}\theta_{it} + \gamma_{2t} \log^2 \theta_{it} + \gamma_{3t} h_{itl} + \gamma_{4t} h_{itl}^2 + \gamma_{5t} e_{itl} + \gamma_{6t} e_{itl}^2 \\ & + \gamma_{7t} \log C_{it} + \gamma_{8t} \log \theta_{it} \times \log C_{it} + \gamma_{9t} h_{it} \times \log C_{it} + \gamma_{10t} e_{it} \times \log C_{it} \\ & + \gamma_{11t} h_{it} \times e_{it} + \gamma_{12t} h_{it} \times \log \theta_{it} + \gamma_{13t} e_{it} \times \log \theta_{it} + \mathbf{z}_{it}' \gamma_{5t} + \nu_{it}.\end{aligned}$$

where h_{it} and e_{it} are parental and child time investment.

[53] Under the technology's parametric assumption, this conclusion can only be reached if the sign is positive. If it is negative the opposite conclusion does not follow. Formally: $\frac{\partial^2 \log \theta_{it+1}}{\partial h_{it} \partial \log C_{it}} = \frac{\partial^2 \theta_{it+1}}{\partial h_{it} \partial C_{it}} \frac{C_{it}}{\theta_{it+1}} - \frac{\partial \theta_{it+1}}{\partial h_{it}} \frac{1}{\theta_{it+1}^2} = \gamma_{9t}$.

Thus, it could be that $\gamma_{9t} < 0$ and $\partial \theta_{it+1m} / \partial h_{it} \partial C_{it} > 0$, that is, there is complementarity in production between inputs even though the sign of the parameter is negative. At this point it is clear how the interpretation of the estimated parameters depends on the assumptions made regarding the technology is parametric functional form. For example, if the function were assumed to be linear, the parameters are the cross derivative as opposed to the elasticity or semi-elasticity.

equation (1.2.6) have a log-log relationship with skills in the next grade. Hence, under the parametric assumption, the equivalent equation (1.7.14) for these inputs represent an elasticity. A 1 percent increase in classroom inputs increases skills between 0.56 and 0.44 percent.[54] The elasticity of skills or previous skills across grades are between 0.86 and 0.54. Following Cunha and Heckman (2008) these effects are interpreted as skills self-productivity.

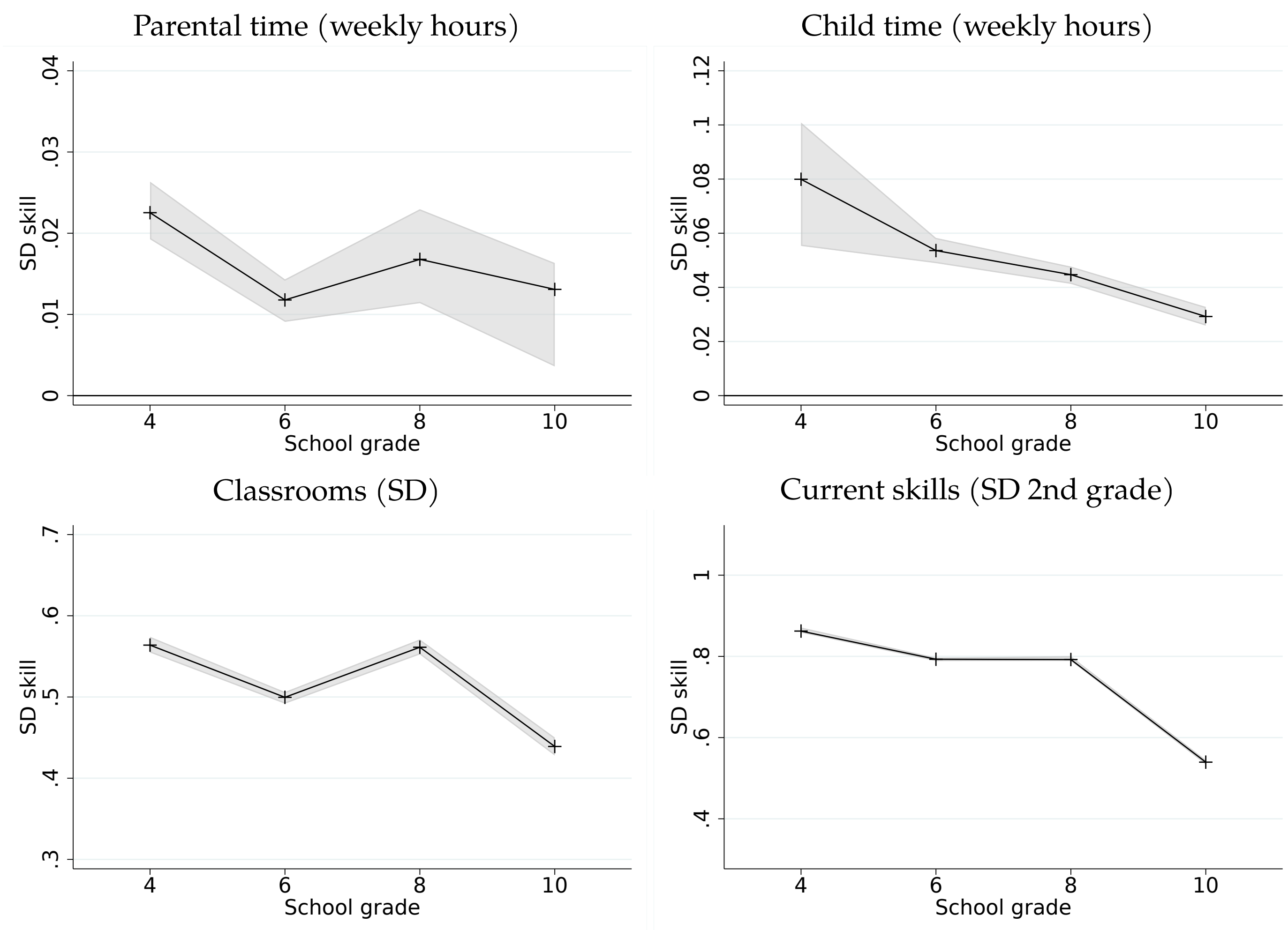
A note of caution applies when comparing these effects across grades. There are two considerations to keep in mind. One, the comparison and conclusions crucially depend on the cardinal normalization of skills across grades, which requires stringent conditions to hold. Second, these are *sample average* marginal effects; decreasing marginal return of inputs could be the reason the average effect of child time investment at tenth grade is lower than at fourth grade.[55] Figure S.1.4 presents the average marginal effect of inputs in terms of changes in percentiles of the skill distribution. The pattern is similar for both time investments. Parental time presents a small decrease after fourth grade; the average effect of one weekly hour goes from 0.4 to 0.3 percentiles. The average effect of one weekly hour is larger for child self-investment; it decreases monotonically from 1.5 to 0.6 percentiles from grades 4 to 10. Classroom effects in terms of percentiles have a U-shape across grades, with values between 10.5 and 9.0 percentiles. Lastly, the average effect of current skills monotonically decreases from over 14 to 12 percentiles.[56]

[54]These effects seem large relative to the literature. The education production function literature finds that a 1 SD increase in the distribution of classroom value-added is associated with an effect on educational achievement between 0.3 and 0.4 SD. These results are larger for two reasons: 1) I am anchoring the skills scale to the standard deviation of skills at second grade. The skills' standard deviations at grades 4, 6 and 8 are around 20 percent larger than that of grade 2, while second and tenth grades' SD are similar (see Table S.1.3). 2) the specifications in Tables S.1.5 and S.1.6 are more general than typical education production functions; they incorporate heterogeneous effects of classrooms. However, standard classroom and teacher value-added models estimated in this data produce results similar to those in the literature. I present these results in Table S.1.7.

[55]Table 1.3 shows the average level of time investment by school grade. Child time investment increases substantially between fourth and tenth grade. At least partially, this drives the decreasing average marginal effect across grades. Note that under decreasing effects across grades, the fact that investment level is increasing implies that the associated cost should decrease relatively more to rationalize the data.

[56]The equivalent results for the specification with within-classroom inputs are similar. These results are presented in Figures S.1.3 and S.1.5.

Figure 1.2:
Sample average marginal effects of skill inputs (SD)



Note: The values on these graphs show the average marginal effect calculated using the estimates from the specifications in Table S.1.5. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student's marginal effect using each input's analogous equation (1.7.14) and calculate the average over the sample.

The technology estimates could be biased if there are unobservable inputs or characteristics that systematically influence skills and inputs. If conditional on the classroom assignment there is selection bias in the estimates of the observable inputs' effects, and if the classroom effects are bias due to omitted variable bias.

The identification assumption relies on the rich data available on inputs and students' characteristics on an attempt to include a sufficient number of key factors that influence the skill formation process. This mean independence assumption is fundamentally untestable. Nevertheless, I follow Chetty et al. (2014a) and Agostinelli et al. (2020) and provide an indirect test of selection in unobservables and test the out-of-sample prediction performance of the estimated technology. The tests require an observable variable that is highly correlated with skills but that is not included in the specification of the skill technology. As in the mentioned studies, I use household income as the omitted observable variable. Since household income is correlated with skills, it is likely correlated with its inputs. Then, if there were important omitted inputs they should be correlated with income. The indirect test of the omitted inputs has as null hypothesis that household income is not correlated with the residual of the skill technology. If the hypothesis is not rejected it suggests that there are not relevant omitted inputs.

Table 1.4 presents the results of this test. The first row shows the coefficients from a regression of skills at each grade on household income—measured in units of 3,000 dollars, which represents around 30 percent of the standard deviation and average annual household income in the sample.[57] An additional 3,000 dollars is associated with skills between 0.12 and 0.18 SD higher. The second row of the table shows the coefficient of regressing the estimated residuals of the skill technology on household income. In this case, an additional 3,000 dollars in household income is associated with residuals between 0.0007 and 0.0037 SD (of skills) higher. These estimates are statistically significant. However, given that a large income change implies only a small associated change in residuals, the

[57] Table 1.1 presents the average monthly household income in Chilean pesos. The equivalent yearly household income in 2018 dollars has an average and standard deviation around 11,000 dollars.

Table 1.4:
Validation Test: Selection on observables — Household income

	Fourth grade (1)	Sixth grade (2)	Eighth grade (3)	Tenth grade (4)
$\log \theta_{it+1}$ (skills)	0.1352	0.1531	0.1824	0.1266
	(0.0043)	(0.0045)	(0.0055)	(0.0042)
	[0.1271;0.1436]	[0.1445;0.1621]	[0.1721;0.1937]	[0.1187;0.1353]
$\hat{\nu}_{it}$ (residual skill tech.)	0.0033	0.0028	0.0037	0.0007
	(0.0004)	(0.0002)	(0.0003)	(0.0002)
	[0.0025;0.004]	[0.0023;0.0032]	[0.0031;0.0041]	[0.0001;0.0011]
N	407,720	596,617	457,782	336,470

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. Skills are $\log \theta_{it+1}$ and $\hat{\nu}_{it}$ represents the residual of the estimated technology from the specification in Table S.1.5. The first row shows the coefficients from a regressing skills on household income, measured in units of $3000 dollars US in 2018 values (0.3 SD of the household annual income distribution). The second row shows the coefficients of regressing the technology's residuals on household income.

results suggest this relationship is not economically significant. The conclusion I draw from the results is that if there is an omitted factor, the implied bias should be minimal.

As in Agostinelli et al. (2020), I use household income to test the *out-of-sample* predictive performance of the estimated skill technology. In particular, I evaluate if the technology is able to predict the average skills by household income deciles. Figure S.1.8 plots the average predicted skills (grey bars) against the average skills observed in the data (white bars) for each decile of the household income distribution and by school grade. The figure shows that the estimated technology predicts the average skill of each household income decile extremely well even though this variable in not included in the specification.

Lastly, classroom effects are biased if the assignment of students to classroom is based on unobserbable factors. To test this hypothesis, I regress the estimated classroom effects on household income. The results are in the first row of Table 1.5. An additional 3,000 dollars is associated with classroom effects between 0.04 and 0.09 SD larger. This suggest that students are assigned to classrooms based on characteristics correlated with income. The second row of Table 1.5 shows the coefficient of the same regression of classroom effects on income but including school fixed effects. In this case the coefficients are below 0.01 SD

Table 1.5:
Coefficient of regression of classroom effects on household income

	Fourth grade (1)	Sixth grade (2)	Eighth grade (3)	Tenth grade (4)
No school FE	0.0493	0.0704	0.0906	0.0896
	(0.0039)	(0.0028)	(0.0037)	(0.0007)
	[0.0418;0.0573]	[0.0648;0.0764]	[0.0812;0.0975]	[0.0840;0.0967]
School FE	0.0080	-0.0035	0.0039	0.0083
	(0.0007)	(0.0008)	(0.0011)	(0.0008)
	[0.0067;0.0093]	[-0.0052;-0.0021]	[0.0027;0.0067]	[0.0068;0.0098]
N	407,720	596,617	457,782	336,470

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. The table presents the coefficients from regressing classroom effects on household income, measured in units of $3000 dollars US in 2018 values (0.3 SD of the household annual income distribution). The first row "No school FE" does not include school fixed effects, while the second row "School FE" adds school fixed effects. The classroom effects correspond to the estimated technology from the specification in Table S.1.5.

and the relationship is not economically significant. This result suggest that most of the selection is through the school choice.[58] This results motivates the estimation strategy of the time investment policy function. The unconditional correlation between time investment and classroom effects would potentially be different than the causal response of parents and students to classroom quality. Including additional time-varying covariates and a student idiosyncratic component in equation (1.6.10) aims to address this selection.

1.7.3 Time investment functions

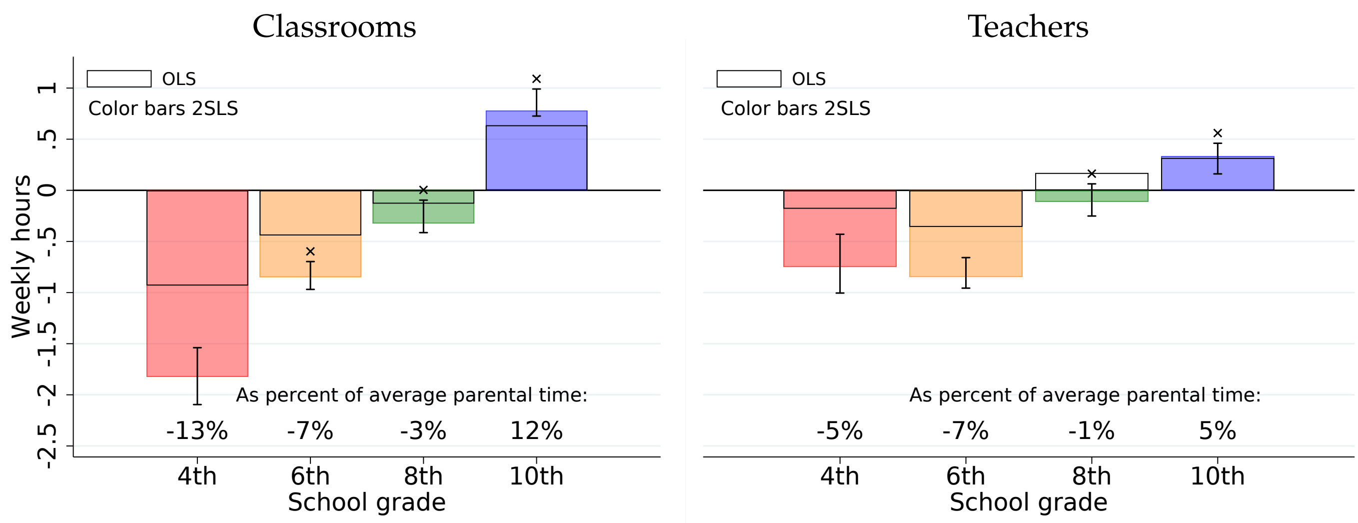
This section describes the estimates of the time investment functions in equation (1.2.9). There are two different types of household time investments: parental and child time investment. Table S.1.8 presents the estimates; columns (1) and (2) show the estimates of parental and child time investment responses to classroom effects associated with the skill technology specification of Table S.1.5, while columns (3) and (4) show the responses to teacher effects associated with the estimated technology reported in Table S.1.6. Note

[58] In Chetty et al. (2014a)'s environment most of the selection is due to school choice as well, rather than across teachers.

that both classroom and teacher effects are normalized to be mean zero and variance one. Thus, the coefficients are interpreted as the response—in weekly hours— to a change of one standard deviation (SD) of classroom or teacher effects.

The left panels of Figures 1.3 and 1.4 present the responses of parents and students to a reassignment of the student from a classroom in the 25th to one in the 75th percentile of the classroom quality distribution, while the right panels show their responses of reassigning the class' teacher from the 25th to 75th percentile of the teacher quality distribution. The empty bars show the estimates without measurement error correction—i.e., the OLS estimates. The colored bars show the estimates using 2SLS to correct for measurement error. The vertical lines on top of each bar represent 95% confidence intervals and the symbol **x** on top indicates the response is statistically significant different at 1% from the fourth grade's response.

Figure 1.3:
Parental time responses (weekly hours)
to reassignment from 25th to 75th percentile of classroom and teacher quality distributions

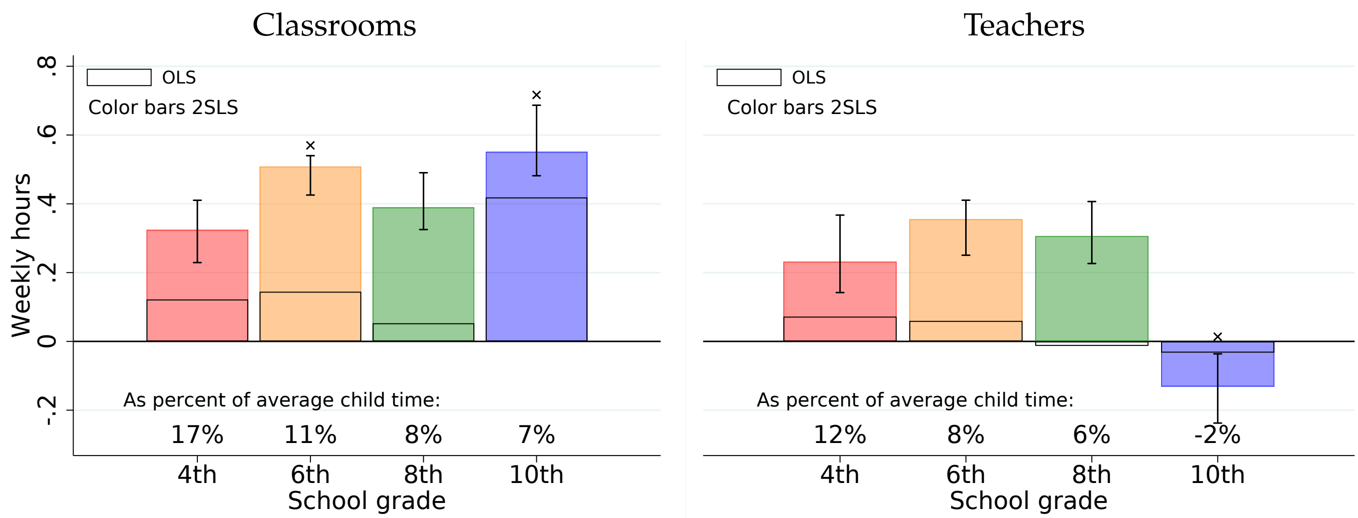


Note: The values on the graphs are calculated using the estimates of Table S.1.8. The vertical lines are school network-clustered bootstrapped 95% confidence intervals. The symbol on top (x) indicates that the difference between the response with that of fourth grade is statistically significant at 1%. The values on the bottom of each plot are the responses as a percent of the average parental time at each grade.

The estimates show that the responses are not homogeneous across school grades.

Parents of fourth graders compensate for the classroom reassignment; they respond by decreasing the time they spend with their children by around 1.8 weekly hours. However, the magnitude of the responses decreases as children grow up. At grade 10, the responses are in the opposite direction, the additional classroom inputs increase parental time by 45 minutes per week. The results for younger children are consistent with the literature of parental responses to specific school inputs (Houtenville and Conway, 2008; Das et al., 2013; Fu and Mehta, 2018). These responses represent 13 and 12 percent of the average parental time in fourth and tenth grades respectively.[59] The responses to teachers follow the same pattern, but with smaller magnitudes. In particular, as a result of the improvement in teacher assignment parental time decreases by 0.7 and 0.8 hours per week for fourth and sixth graders and in tenth grade parental time increase by a quarter hour per week.

Figure 1.4:
Child time responses (weekly hours)
to reassignment from 25th to 75th percentile of classroom and teacher quality distributions



Note: The values on the graphs are calculated using the estimates of Table S.1.8. The vertical lines are school network-clustered bootstrapped 95% confidence intervals. The symbol on top (x) indicates that the difference between the response with that of fourth grade is statistically significant at 1%. The values on the bottom of each plot are the responses as a percent of the average child time self-investment at each grade.

As a response to the same classroom reassignment students increase time investment at every grade; by 20 minutes per week in fourth grade and by half an hour per week

[59]The average time investments by school grade are presented in Table 1.3.

in the upper grades (Figure 1.4). These responses are large relative to the average child self-investment (between 17 and 7 percent, respectively).[60] Meanwhile, the responses to teachers are smaller as for parents and the response is virtually zero for tenth graders. One possible reason teacher quality induce no response in tenth grade is that teachers in lower grades tend to teach most subjects and so they spend more time interacting with students.[61]

These responses imply a non-trivial impact on skills. An increase of 1 SD in classroom effects implies a response of fourth graders' parents that decreases skills in 0.036 standard deviations (SD) of log skills in second grade. At tenth grade the parental response increases skills by 0.022 SD. Meanwhile, students' responses increase skills in 0.044 and 0.023 SD in grades 4 and 10, respectively. The impact of the responses on skills represent between 3 and 11 percent of the overall effect of classrooms on children.[62]

It is difficult to pinpoint the reasons of the change in the direction for the parental responses to school inputs. In the model this is the result of the interaction of the evolution of the skill formation technology as children grow up and preferences or costs associated with investments at different points in the development of children. The technology might vary across education stages for many reasons. As an example, parents and teachers might be more "substitutable" to teach certain skills when children are younger, such as basic math. As children start learning subjects with more specialized knowledge—e.g., calculus—parents are less able to substitute for teachers. However, parents might spend relative more time with their children in different activities, like advising, that could be more complementary to teacher quality in the formation of skills.[63]

[60]Figure S.1.10 shows the responses to classroom and teacher reassignment between percentiles 10 and 90 across each distribution. The magnitudes can be quite large, e.g., a decrease of almost 3.5 weekly hours in grade 4, and 1.5 weekly hours increase in tenth grade. Meanwhile, for students the figure shows an increase of student effort of over half a weekly hour in fourth grade and almost a weekly hour in the remaining grades.

[61]Students interact, on average, with around 6 teachers in fourth grade and with almost 11 teachers in tenth grade (Table 1.1). Moreover, usually in the lower grades there is a main teacher who is in charge of the "core" subjects, such as math, language, biology, and so on, while auxiliary teachers teach music, art and physical education.

[62]The impact on skills of parents' responses are -0.036, -0.008, -0.005 and 0.010 SD and the effects of children's responses are 0.024, 0.021, 0.011 and 0.010 SD for grades 4, 6, 8 and 10, respectively

[63]Figure S.1.9 presents results from the National Household Education Survey that attempt to provide suggestive evidence regarding this conjecture. It shows the share of parents that help with homework at

1.8 Conclusions

In this chapter, I study the time investment responses of students and their parents to the quality of school inputs and how these responses evolve as children grow up. I combine administrative and survey data from Chile to estimate the household responses to classroom inputs and teachers from grades 4 to 10. The responses differ by school grade; parents of fourth graders compensate for classroom and teacher quality, while parents of secondary school students reinforce quality. Students increase effort if the classroom environment improves, with larger magnitudes for older children. Moreover, household responses to teachers have smaller magnitudes but show a similar pattern across grades. However, students virtually do not adjust their time investments for different teacher quality at grade 10.

The estimates shed light on the mechanics at play in the black box of classroom and teacher value-added. Further, they inform policy design on teacher selection and pay-for-performance.[64] Simulations of policies that remove the lowest-performing teachers (e.g., Hanushek, 2011; Goldhaber and Theobald, 2013; Chetty et al., 2014b) and analysis of optimal rules for teacher dismissal (Staiger and Rockoff, 2010; Neal, 2011) could be improved by addressing the behavioral response of households. For example, these exercises usually include students at different school grades. If behavioral responses vary by students' age, then these policies hold some teachers to a higher standard than others, which generates inefficiency costs. Similarly, families' responses might weaken the link between rewards and teacher effort, leading to ineffective policy schemes (Neal, 2011).[65] Behrman et al. (2015) study teacher-incentive schemes and find no effect on academic achievement, unless the

least 3 days in an average week and the share of parents that discussed time management in the past week across ages of children. The former shows a monotonic decline while the latter slightly increases as children grow up. This result suggests that parents might modify the composition of the activities considered as time investments as children grow up.

[64] See Jackson et al. (2014) for a review of the literature on teacher-related policies.

[65] As an example, Springer et al. (2011) find that the teacher-incentive POINT program in Tennessee did not result in performance improvements of students assigned to eligible teachers. Neal (2011) argues this may be explained by too high performance targets. This is exacerbated if teacher effort *crowds out* parental investment, dampening the impact of the policy.

program includes an additional student-incentive component. Their result could be driven in part by a decrease in parental investment in response to teacher effort. The estimates of household responses suggest potential gains from schemes that include a parent-incentive component, in addition to students' and teachers' components.[66] Moreover, the Chilean government introduced a teaching reform that sets new teacher hires to be compensated based on measures of competency. Not taking into account the heterogeneous household responses could lead to inefficiency costs in the implementation of the policy.[67]

[66] Levitt et al. (2016) study an incentive-based program with students and parents as beneficiaries, but it did not include a teacher-incentive component.

[67] The government introduced the policy in 2017 and it is gradually implemented through 2023. See Tincani (2020) for a detailed description of the policy and an ex ante evaluation of the Chilean merit-based teaching reform.

Chapter 2

Optimal Allocation of School Resources Across Grades

Chapter summary

The heterogeneous responses by grade found in chapter 1 motivate the analysis of optimal resource allocation policies across education levels. Chapter 2 builds on chapter 1 to understand how the differential impact by grade of school resources and home investment can be used to design the optimal allocation of school resources across grades. To that end, I build and estimate a child development model using an indirect inference approach. I use the estimated model to simulate counterfactuals of the dynamics of the cognitive skills of students and I characterize the optimal allocation of school resources across grades. The results suggest that, on average, it is optimal to allocate relatively more resources in lower grades than in upper grades with respect to the allocation observed in the data. Moreover, the behavioral response of households plays a key role in the characterization of the optimal allocation.

2.1 Introduction

A dynamic skill formation technology and the differential response of households to classroom quality by grade found in chapter 1 raises the question: what would be the optimal allocation of school resources, such as teachers or monetary resources, *across* school grades? For example, a teacher's direct impact on cognitive skills and parents' and children's time investment responses depend on the school grade. The optimal assignment of the teacher should account for these differential effects across grades. The estimates of the household responses inform how parents and children adjust their time investments based on *exogenous* changes in classroom inputs. Unless households are myopic, these responses vary with changes in the expected classroom environments of subsequent grades. To evaluate the implications of different allocations of resources across grades it is necessary to add structure on preferences and on the expectations process of households. To that end, I build and estimate a dynamic child development model based on a unitary household that maximizes lifetime utility by choosing parental and child time investment subject to the skill formation technology.

I estimate the child development model with an indirect inference estimator. The auxiliary model consists of the time investment policy functions and moments of the conditional distribution of skills and time investment. The child development model then allows me to characterize the allocation of resources across grades that maximizes each student's cognitive development. I find that the optimal allocation improves skills by 0.20 standard deviations (SD) with respect to the baseline allocation. On average, it is optimal for schools to invest relatively more resources in lower grades. Moreover, the behavioral response plays a key role in characterizing the optimal allocation. Ignoring the response of households to classroom quality leads to an optimal allocation that yields cognitive improvements that are between 25 and 65 percent lower. These findings highlight the importance of the role of the behavioral response in school resource allocation policies.

This chapter contributes by characterizing the optimal allocation of school resources across grades. There is a line of research that studies the dynamic complementarity of investments in the formation of cognitive skills (e.g., Cunha et al., 2010) and Johnson and Jackson (2019) show evidence of complementarity of school investments at different education levels. However, the literature does not address how school resources should be distributed across school grades. Furthermore, the heterogeneous response of households adds an additional margin to optimize the allocation of school resources. Using the child development model that explicitly incorporates school inputs, I characterize the allocation that maximizes cognitive development while considering the differential impact and relationship of inputs across grades in the skill formation process as well as the behavioral response of households.

The structure of the chapter is as follows. In Section 2.2, I specify the parameterization of preferences and the expectations process, outline the estimation methodology of the child development model and describe the estimates. Section 2.3 consists of the policy counterfactual analysis of the optimal resource allocation across grades. Finally, Section 2.4 concludes the chapter.

2.2 Model parameterization and estimation

The estimates in the previous chapter are the responses of parents and students to classroom and teacher quality holding everything else constant. The policy space that these estimates are able to evaluate is constrained to policies that do not change the expected classroom environments of subsequent grades. For example, consider evaluating the impact of increasing fourth grade's resources at the expense of reducing resources at grade 10. If we were to evaluate this policy with the estimated skill technology and the time investment functions, we would assume that households respond to the additional resources in fourth grade ignoring any resource changes at tenth grade. By the time students attend grade 10, they receive lower resources at school and respond accordingly. If households were aware of the resources reallocation across grades, these results do not represent the true policy's impact.

In this section, I parameterize the preferences of the child development model described in Section 1.2 of Chapter 1. The model allows evaluation of the policies that require households to update expectations of future classroom environments. I use the model to evaluate different allocations of resources across grades and characterize the allocation that maximizes the cognitive development of children.

2.2.1 Model parametrization

I depart from the standard utility function in terms of consumption. In the model, households value the cognitive skills of the child and incurs in a disutility cost for each hour of parental and child time investment.[1] I relabel parental and child time investment as h_{it} and e_{it}, respectively. The utility functional form is:

$$u_{it}(\theta_{it}, h_{it}, e_{it}, \mathbf{x}_{it}) = \frac{\theta_{it}^{1-\phi_{1t}} - 1}{1 - \phi_{1t}} - \phi_{2it} h_{it} - \phi_{3it} e_{it} \tag{2.2.1}$$

where ϕ_{1t} is the curvature parameter on skills and $\phi_{2it} > 0$ and $\phi_{3it} > 0$ are the disutility costs of parental and child time investment, respectively. The disutility cost parameters are indexed by i to allow for heterogeneity in preferences across households. In particular, $\phi_{2it} = \exp(\tilde{\phi}'_{2t}\mathbf{x}_{it} + \upsilon_i)$ and $\phi_{3it} = \exp(\tilde{\phi}'_{3t}\mathbf{x}_{it} + \iota_i)$, where $\upsilon_i \sim N(0, \sigma_\upsilon^2)$ and $\iota_i \sim N(0, \sigma_\iota^2)$. The vector x_{it} consists of demographic characteristics. Note that the disutility costs are correlated with classroom effects through x_{it}. This captures the observed correlation between classroom effects and investments.

All parameters are index by t. The age-specific parameters associated with children's disutility cost follow from existing work on child development.[2] However, it is less common

[1] A more natural modeling assumption is for the utility to depend on time inputs through forgone consumption. However, since I do not observe leisure or the hourly wage, I cannot adopt that specification.

[2] For example, Del Boca et al. (2019) assign age-dependent discount rates for children following the work by developmental psychologists that shows that the capacity to delay gratification changes substantially as

for parental time's disutility cost to vary across children's age. The age-specific parameter implies that parental time is a different consumption good (bad) depending on the age of the child.

The household has time constraints for both types of investment, $h_t \in [0, H]$ and $e_t \in [0, E]$, where H and E are the time endowments of parents and children, respectively. However, this is not the usual time constraint of economic models where households allocate the time endowment between leisure and hours of work. Instead, this constraint represents an interval of the possible choices of time investment. The state space is given by $\Omega_{it} = \{\theta_{it}, C_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\}$. Where C_{it} is classroom inputs, the vector $\mathbf{z}_{it}$ include demographic characteristics that influence total factor productivity in the production of cognitive skills. The vectors $\mathbf{z}_{it}$ and x_{it} include elements which are grade-invariant like parental education and elements like parent's age that deterministically evolve across grades. Additionally, x_{it} includes household income y_{it} which is modeled as an AR(1) random process.

The cognitive skills of a child θ_{it} follow a first order Markov process, similar to Cunha and Heckman (2008), Cunha et al. (2010), Agostinelli and Wiswall (2016), and Agostinelli et al. (2020):

$$\theta_{it+1} = F_t(\theta_{it}, h_{it}, C_{it}, \mathbf{z}_{it}, \nu_{it}), \tag{2.2.2}$$

where $F_t(\cdot)$ is a grade-specific function that depends on current skill, θ_{it}, the time the household invest in its child h_{it}, the classroom inputs C_{it}, household characteristics $\mathbf{z}_{it}$, and the structural shock ν_{it}.

I assume the technology is a trans-log production function. However, the relationship between the logarithms of skills and time investment is linear, allowing for a corner solution.[3] This parametrization is flexible in terms of the relationship between inputs in the production of skills—i.e., it is possible to have a negative cross-derivative (see Agostinelli and

children grow up (e.g., Steinberg et al., 2009). Moreover, studies show that attention capacity varies across children at different development stages (e.g., Rueda et al., 2004)

[3] In the data, I observe a non-trivial fraction of households choosing to invest zero time in the skill formation process.

Wiswall, 2016). It allows complementarity or substitutability between time and classroom inputs.[4] The parametric functional form is the following:

$$\begin{aligned}\log\theta_{it+1} &= \log F_t(\theta_{it}, h_{it}, C_{it}, \mathbf{z}_{it}, \nu_{it}) \\ &= \log A_t + \gamma_{1t}\log\theta_{it} + \gamma_{2t}h_{it} + \gamma_{3t}\log C_{it} + \gamma_{4t}h_{it}\times\log C_{it} + \mathbf{z}_{it}'\gamma_{5t} + \nu_{it},\end{aligned} \tag{2.2.3}$$

where $A_t\exp(\mathbf{z}_{it}'\gamma_{5t})$ is the total factor productivity. The set $\{\gamma_{jt}\}_{j=1}^5$ defines the elasticity or semi-elasticity of next grade's skills and inputs and ν_{it} is a mean zero shock.

The specification is more general in the empirical implementation. Besides the terms in equation (2.2.3), it includes second order polynomials of current skill and time inputs, interactions between time investment, classroom inputs, and current skills and interactions between time investments of different members of the household—i.e., parental and child time investments. However, equation (2.2.3) reduces notation burden and the identification and estimation analysis under this simplification is without loss of generality.

The household has rational expectations and forms expectations over three variables: future classroom inputs, skill shocks and future household income shocks. I assume the classroom process is:

$$\log C_{it} = \kappa_t'\mathbf{x}_{it} + \Delta_{it}, \tag{2.2.4}$$

where κ_t are grade-specific parameters and $\Delta_{it}\sim N(0,\sigma^2_{\Delta,t})$. I assume the distribution of skill shocks ν_{it} is $N(0,\sigma^2_{\nu,t})$. Finally, the household income AR(1) process is:

$$\log y_{it} = \overline{y}_t + \rho_t\log y_{it-1} + \omega_{it}, \tag{2.2.5}$$

where $\{\overline{y}_t,\rho_t\}_t$ are parameters and $\omega_{it}\sim N(0,\sigma^2_{\omega,t})$. The timing of the model is as follows: First, ν_{it-1}, Δ_{it}, and ω_{it} are realized at the beginning of grade t, and the household learns Ω_{it}. Next, the household makes the time investment decisions without knowing ν_{it}.

[4] Assuming a Cobb-Douglas or constant returns to scale production function with standard parameters values implies weakly complementarity between all the inputs. It is important to allow for this flexibility since the signs of cross-derivatives are relevant in terms of households' responses.

Finally, the terminal value at age T (grade 10) is defined as:

$$V_T(\Omega_{i,T}) = \phi_4 \frac{\theta_{iT}^{1-\phi_{1T}} - 1}{1 - \phi_{1T}} \tag{2.2.6}$$

where $\phi_4 > 0$. The terminal value can be thought as the initial condition of a new problem.

2.2.2 Model estimation procedure

I implement a two-step estimation procedure to reduce computational burden. In the first step, I estimate the measurement system, skill formation technology and household income and classroom processes. The estimation methodologies of the measurement systems and skill technology are described in Sections 1.6.1 and 1.6.2 of Chapter 1, respectively. I estimate the household income and the classroom processes defined in equation (2.2.5) and (2.2.4) with the OLS estimator.

The second step estimates the preference parameters. This step is an indirect inference estimator. Let M be the set of targeted moments which I describe below. The estimator is a simulation-based method. Given the primitive preference parameters Σ and initial conditions, I simulate the households' choices and skills across school grades.[5] I then compute the analogous targeted moments in the simulated data, denoted by $M_S(\Sigma)$, and the estimator is defined as:

$$\hat{\Sigma} \equiv \underset{\Sigma}{\text{argmin}} \; (M - M_S(\Sigma))' \times W \times (M - M_S(\Sigma)), \tag{2.2.7}$$

where W is a weighting matrix. I set the weighting matrix to be the inverse of the diagonal variance–covariance matrix of the moments computed by bootstrapping the data. The auxiliary model consists of: (i) the time investment functions (ii) first- and second-order moments for the conditional distribution of skills and time investments.

[5] Appendix D describes the solution of the model.

2.2.3 Model estimates

The estimates of the measurement system and skill technology are presented and described in Sections 1.7.1 and 1.7.2 of Chapter 1, respectively. Additionally, Table S.2.9 provides the estimates of the variance of the skill technology shock at each school grade. Table S.2.10 and S.2.11 present the estimates of the household income and classroom processes. Finally, the estimates of the indirect inference estimator of the second step appear in Table S.2.12. That is, the parameters of the curvature of the utility with respect to skills and the parameters governing the disutility of parental and child time investment.

2.2.4 Model fit

Figure 2.1 compares the values of the data moments with moments from the simulated data. There are four different sets of moments, the moments of the auxiliary model given by the policy functions of parental time and child effort as well as the expected value and the variance of the conditional distribution of skills.[6]

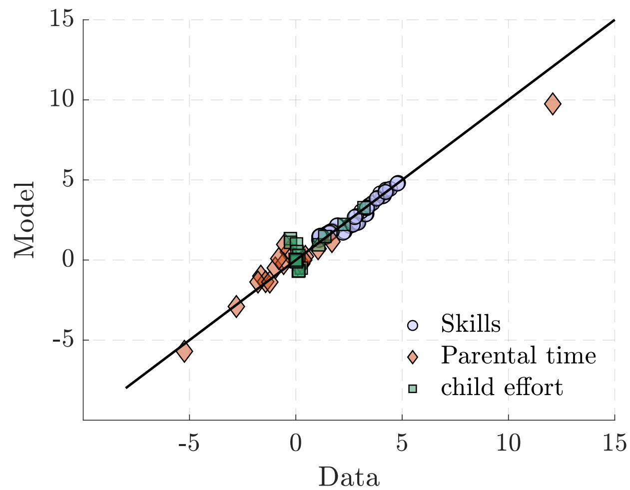
Additionally, Figure 2.2 shows the model fit in terms of the average and standard deviation of cognitive skills, parental and child time investment. The model estimates are able to predict the average skill and investments across grades. The model also predicts reasonably well the standard deviation of the skill, and somewhat larger standard deviation for time investment.

2.3 Optimal school resource allocation policies

The dynamics of the skill formation technology and the heterogeneous responses of households across school grades suggest that, depending on how we distribute the available resources across school grades, we can influence the skill accumulation paths of children. For example, additional resources at grade 10 have a direct effect on students' cognitive

[6] Figure S.2.12 expands the graph for each group of moments.

Figure 2.1:
Model fit of targeted moments



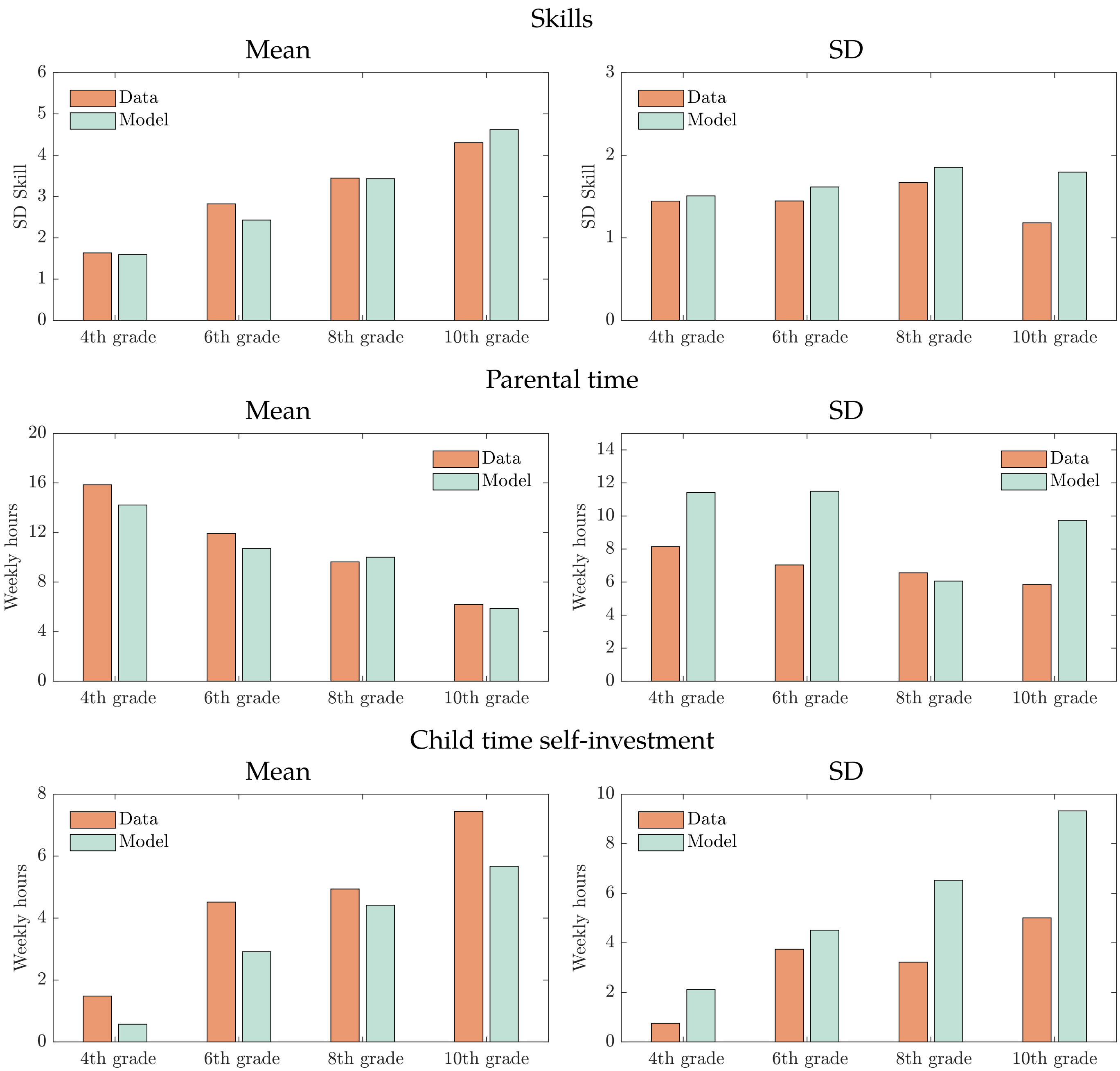
Note: Each dot represents a moment of the auxiliary model in the indirect inference estimator. The horizontal axis shows the values of the moments true data and the vertical axis shows the values of the moments from the simulated data.

skills. However, allocating those resources at earlier grades improves cognitive skills across grades through the self-productivity of skills and so at tenth grade as well. The optimal allocation trade off between the direct effect of resources and the effects through self-productivity of skills. On top on this, households respond differently across grades, which has additional implications for the optimal allocation.

This section presents the policy counterfactual analysis of optimal resource allocation across school grades. The analysis is at the student level—i.e., I evaluate what is the optimal allocation for each student. In this section, I evaluate policies that aim to capture the consequences of the schools' allocation decisions regarding monetary resources, teacher assignment, and other transferable resources across grades.

In addition, optimization requires taking a stand on the objective function. The goal is to evaluate the allocation that maximizes students' well-being. A welfare proxy could be earnings in adulthood, but this information is not available. Another strategy is to maximize skills at the terminal period. However, existing work shows that the effects

Figure 2.2:
Model fit of targeted moments
Mean and SD of skills and parental and child time investment



Note: The bars labeled data and model show the mean and standard deviation (SD) estimated in the real data and in simulated data generated by the child development model.

of teacher and class quality on test scores fade out in subsequent grades but "reemerge" in later outcomes (Deming, 2009; Heckman et al., 2010; Chetty et al., 2011, 2014b). It is possible that skills accumulated in a particular grade are not reflected in subsequent grades' test scores, even though these skills are valued later in life. Thus, setting the terminal skills as the objective function could fail to take into account all gains from certain allocations.

To address this issue, I set as the objective function a weighted average of skills across grades. Each grade weight assigns value to the skills accumulated at that particular grade. I define the weights as follows: For a subset of students in my sample,[7] I observe if they enrolled in college in the year following high school graduation.[8] For these students, I regress an indicator variable of college attendance on the skill measures at each school grade. I then set as the weights the coefficients of this college attendance regression. That is, the (student-specific) optimal policy maximizes the *weighted average of skills across grades* or *weighted skills index*. The weights (normalized to sum up to one) are 0.10, 0.17, 0.24, and 0.49 in grades 4, 6, 8, and 10, respectively.

Lastly, I need to specify households' behavior and their expectation process. I perform the simulations for three household *types* based on different assumptions regarding their behavior: 1) No response: households do not adjust their time investments to different resource allocations; 2) Policy-myopic: households respond to the contemporaneous changes in resources but under the belief that in subsequent grades the resources are those of the baseline allocation; and 3) Forward-looking: household are forward-looking and make their decisions internalizing the dynamic implications of different allocations.

I simulate choices and skills using different estimates for each type. For non-responsive households, I simulate the counterfactual of different allocations using the skill formation technology. If households are policy-myopic, I simulate their choices using the approximated policy functions of time investment. Note that the estimated time investment functions do not take a stand on the expectation process—e.g., if households are myopic or

[7] The subset of students corresponds to students that by 2018 could have graduate from high school

[8] This information is available in the administrative data of higher education of Chile.

forward-looking. However, simulations based on these estimates implicitly assume that households are not taking into consideration that the expected resources in future grades are different than those in the baseline allocation. If household were truly myopic (not only in terms of the policy), the time investment functions provide the correct responses for any policy implemented. Finally, the simulation under forward-looking households is carried out with the dynamic child development model.

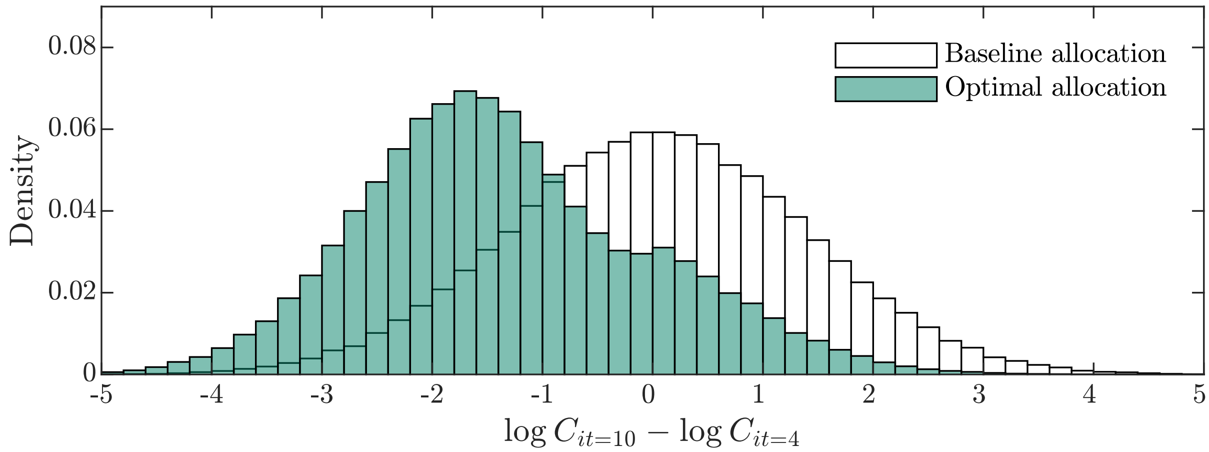
2.3.1 Policy-myopic households

In this exercise, I consider fourth graders in the sample and I simulate their classroom effects at each grade using equation (2.2.4). The time investment functions and the skill technology provide their choices and skills' dynamics, respectively, for every possible assignment across grades of the classroom effects $\{\log C_{it}\}_{t=4,6,8,10}$. The optimal allocation for each student maximizes their weighted average of skills. Note that reallocating classroom effects across grades might seem extreme in terms of feasibility. However, even though this has implications for the magnitude of the policy's impact on skills, given the linearity of the functions, it does not alter the conclusions drawn from the optimal allocation. Additionally, the exercise in done under the implicit assumption of indivisibility of the classroom effect. Nevertheless, in the next section I account for these potential restrictions in the allocation decisions. The current exercise help us understand the mechanics of the impact of different resource allocations.

Figure 2.3 plots the distribution of the difference between the classroom effects in tenth and fourth grades ($\log C_{it=10} - \log C_{it=4}$). Positive values imply school resources in grade 10 are larger than in grade 4 and *vice versa*. Pushing on the technology's parametric assumption, this difference represents the logarithm of the rate of late to early classroom investments.[9] The white bars show the distribution under the baseline allocation and the

[9]This is similar to the exercise implemented in Cunha et al. (2010) regarding the optimal ratio of early to late investments. The main difference is that their optimization problem allows resource reallocation across children, while in the current setting resource reallocation is within students over time.

Figure 2.3:
Policy myopic households: Optimal allocation that maximize (weighted) skills

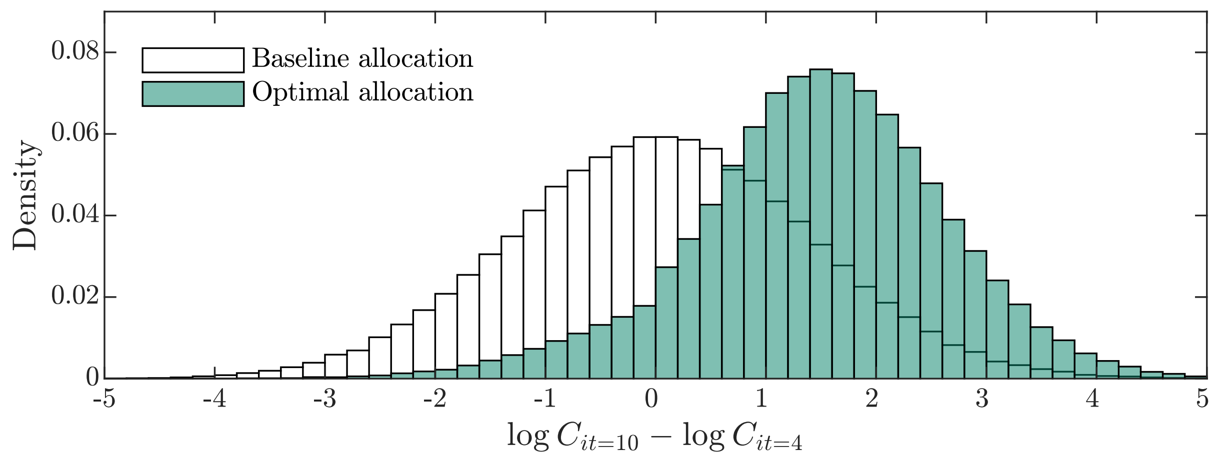


Note: The white bars plot the distribution of the difference between the classroom effects in grades 10 and 4, under the baseline allocation. The colored bar shows the difference between the classroom effect in grade 10 and 4 resulting from the optimal allocation. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the time investment functions (Table S.1.8) and the skill technology (Table S.1.5) I simulate the household choices and their skills for every possible assignment of the realizations of classroom effects $\{\log C_{it}\}_{t=4,6,8,10}$ across grades. The optimal allocation maximizes the weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17,.0.24 and 0.49 for grades 4, 6,8 and 10, respectively. Note that using the investment functions for the simulations implicitly assumes that households respond as if resources in subsequent grades are the given by baseline allocation.

colored bars plot the distribution resulting from the allocation that maximizes the weighted skills index. For most students it is optimal if relatively more resources are allocated in fourth grade than in tenth grade. Moreover, the distribution shifts substantially toward the left suggesting that the baseline allocation is far from "optimal" and that there is room for improvement under a balanced budget policy.

In addition, I perform the optimization setting as the objective function the terminal skills of students. Figure 2.4 shows the distribution of the difference between the classroom effects in tenth and fourth grades with terminal skills as the objective function. In contrast to the case with the weighted skills index as the objective function, the distribution of the difference between *late* and *early* school investments shifts toward the right—i.e., investing

Figure 2.4:
Policy myopic households: Optimal allocation that maximize skills at grade 10 by mother education



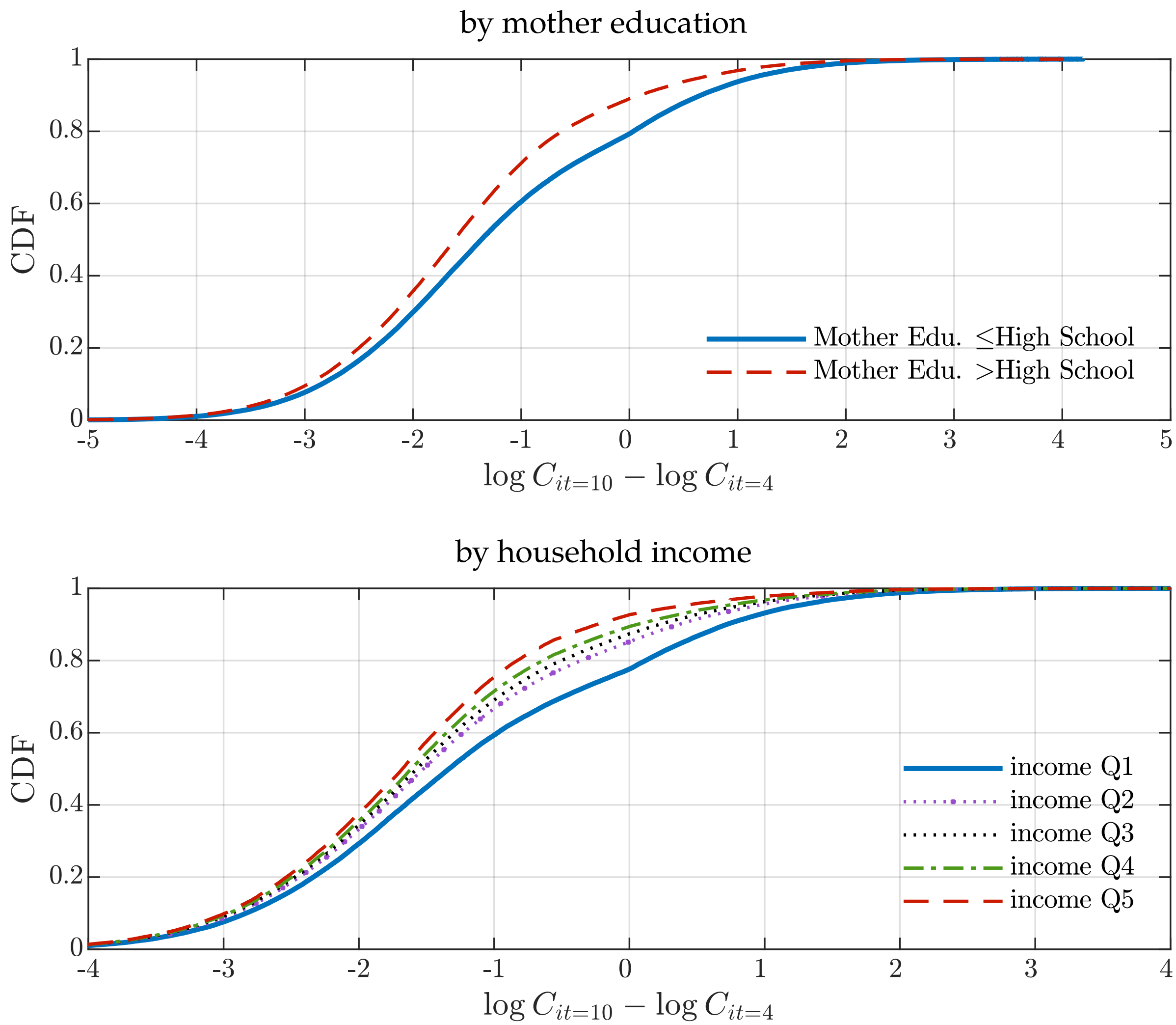
Note: This figure shows the allocation that maximizes the skills at grade 10 of each student, as opposed to maximizing a weighted average of skills. See note in Figure 2.3 for details.

relatively more in the upper grades maximizes the terminal skill stock of children. However, assigning even a small weight to skills in previous grades alters the conclusion, leading to a graph similar to Figure 2.3.

To characterize the optimal allocation of students of different backgrounds, I split the distribution based on the demographic characteristics of students. The two panels of Figures 2.5 and **??** show the distribution of the difference of classroom effects that maximizes the weighted skills index but splitting the sample by the education level of students' mothers and household income quintiles, respectively. The bottom panels show the corresponding cumulative distribution in each case. These graphs show that a larger share of students from more affluent backgrounds have a negative difference between classroom effects. For example, only around 10 percent of students whose mothers have more than high school education have a (optimum) positive difference while this share is 20 percent for students whose mothers have high school or less education. This is largely driven by higher self-productivity of skills for students from more affluent backgrounds.[10]

[10]See Figures S.1.6 and S.1.7, where the sample average of the marginal effects of current skills is larger for students from more affluent families. This result is also invariant to the choice of objective function.

Figure 2.5:
Policy myopic households: Optimal allocation that maximize (weighted) skills



Note: See note in Figure 2.3 for details.

2.3.2 Forward-looking households

I use the child development model to evaluate resource allocations across grades and define the optimal allocation of each student as the one that maximizes their weighted skills index. Relying on the parametric assumptions on the skill formation technology and the classroom process, a student has expected classroom inputs at each grade, conditional on x_{i0}, given by:[11]

$$R_{it} = \exp\left[\int \cdots \int (\kappa_t' \mathbf{x}_{it} + \Delta_{it}) \times f_{\Delta,t}(\Delta_{it}) \times d\Delta_{it} \times \prod_{t' \leq t} f_{\mathbf{x}_{t'}|\mathbf{x}_{it'-1}}(\mathbf{x}_{it'} \mid \mathbf{x}_{it'-1}) \times d\mathbf{x}_{it'}\right], \tag{2.3.8}$$

where $f_{\Delta,t}$ is a density function of a normal distribution with mean zero and variance $\sigma^2_{\Delta,t}$ and $f_{\mathbf{x}_t|\mathbf{x}_{it-1}}(\mathbf{x}_{it} \mid \mathbf{x}_{it-1})$ is the transition function of $\mathbf{x}_{it}$. Thus, we can choose a policy that allocates optimally those resources across grades.

A concern is that schools might face limitations on "moving" resources from one grade to another; the classroom effects include school inputs such as teachers or monetary resources that can be easily reallocated and others, like peer composition, that are not transferable. Calculating the share of transferable resources is a complex task. As a simple solution, I assume only 30 percent of inputs are transferable across grades. Nevertheless, below I show results that suggest that this *ad hoc* assumption does not play an important role. At most, setting the share of transferable resources has small implications for the magnitudes of the impact of different allocations, but it does not affect the qualitative conclusions.

Let s_{it} be the share of total transferable resources assigned to grade t. The policy assigns expected classroom inputs at each grade given by:

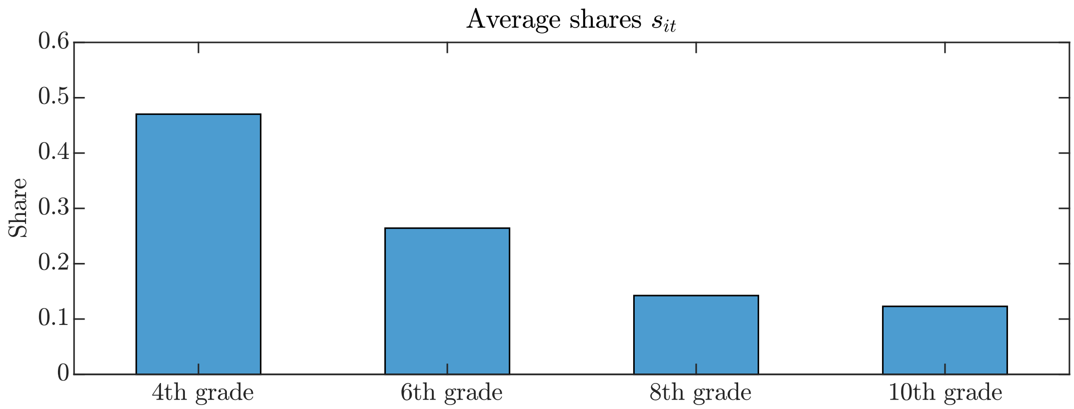
$$\mathrm{E}[C_{it}] = 0.7 \times R_{it} + s_{it} \sum_{t'} 0.3 \times R_{it'}. \tag{2.3.9}$$

That is, the expected classroom inputs at grade t are given by 70 percent of the baseline's resources at grade t plus a share s_{it} of the total transferable resources (the sum of 30 percent of baseline resources at each grade). The optimal shares $\mathbf{s}_i^* \equiv \{s_{it}^*\}_t$ are defined as:

$$\begin{aligned} \mathbf{s}_i^* \quad \equiv \quad & \underset{\mathbf{s}_i}{\operatorname{argmax}} \quad \textstyle\sum_t w_t \times \log \theta_{it} \\ & \qquad\qquad \text{subject to } \textstyle\sum_t s_{it} = 1, \end{aligned} \tag{2.3.10}$$

[11]The law of motion of classroom effects is $\log C_{it} = \kappa_t' \mathbf{x}_{it} + \Delta_{it}$. Before attending grade t, there is uncertainty about the realization of the classroom shocks Δ_{it} and the income process shock ω_{it} (note that x_{it} includes household income).

Figure 2.6:
Optimal resource allocation across grades
Average shares of transferable resources



Note: This figure shows the average optimal shares of transferable resources at each school grade. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the child development model, I simulate the household choices and skills under each possible resource allocation across grades for each student. The optimal allocation is given by the shares of total transferable resources at each grade that maximize the weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17, 0.24 and 0.49 for grades 4, 6, 8 and 10, respectively. In this exercise I allow 30 percent of the resources of the baseline allocation of each grade to be transferable across grades. See section 2.3.2 for additional details.

where the weights w_t are the coefficients from the regression of college attendance on skill measures.

Figure 2.6 shows the average (across students) optimal shares of transferable resources assigned at each school grade.[12] On average, the shares are decreasing across grades; at fourth grade is around 47 percent and it decreases monotonically to 12 percent in grade 10. Figure S.2.13 shows the average optimal shares by mother education and household income quintiles. Similar to the policy-myopic setting, students from more affluent backgrounds, given their higher skills' self-productivity, tend to be assigned a larger share of transferable resources at the lower grades.

Figure 2.7 shows the difference in average cognitive skills, classroom effects and time investment between the optimal and baseline resource allocations. The difference in skills

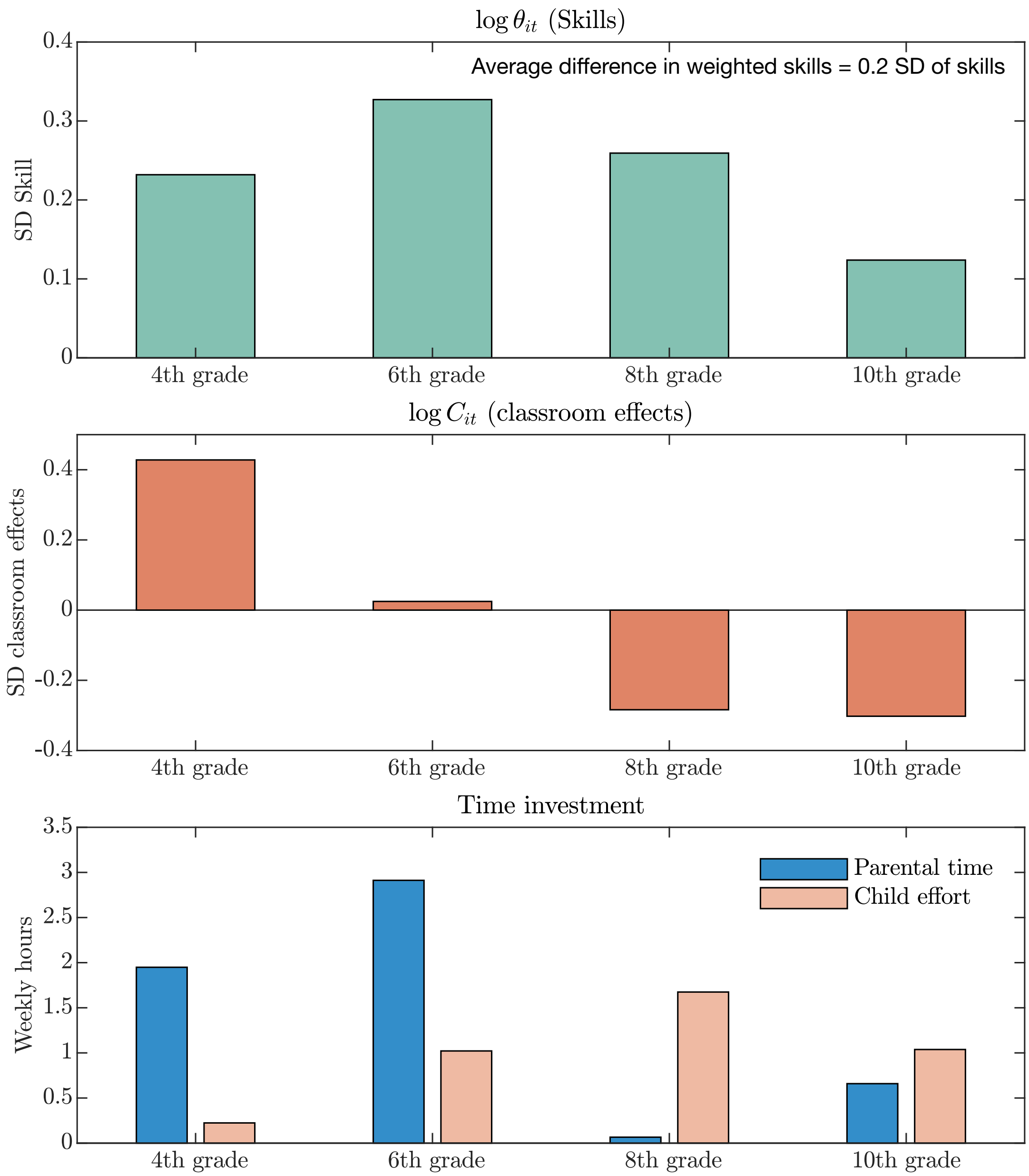
[12] Figure S.2.17 presents the distribution of the shares by school grade.

follows a inverse-U shape across grades and represents an increase of the weighted skills index of 0.2 SD. In grade 4, on average, students receive additional school resources relative to the baseline, which leads to an improvement in their cognitive skills. Households update their expectations about the future school environment and increase parental and child time investment. This improves cognitive skills in addition to the direct effect of the class inputs. In subsequent grades, skills increase because the skills at the start of the grade are, on average, higher. At these grades, changes in classroom inputs and parents' and children's responses affect cognitive skills as well. However, in grade 6, average classroom effects are practically unchanged and in grades 8 and 10 there are fewer resources with respect to the baseline allocation, leading to an improvement in cognitive skills relatively smaller than that in grade 6.

In order to understand the relevance of the *ad hoc* assumption of 30 percent transferable resources at each grade, I carry out the same exercise under different constraints. The top panel of Figure S.2.14 shows the average increase in the weighted skills index of implementing the optimal allocation allowing from 10 to 50 percent of the baseline resources at each grade to be transferable. If only 10 percent of the resources are transferable, this is enough to increase the weighted skills index by 0.15 SD—25 percent lower than in the case of 30 percent transferable resources. As we increase this percentage, the set of possible allocations expands. Changing the restriction from 30 to 40 percent implies virtually zero change in the weighted skills index. That is, for most household this constraint is not binding and the unconstrained optimal allocation is (or is almost) achieved.

In addition, the bottom panel of Figure S.2.14 shows the policy impact on the average weighted skills index for the top and bottom 20 percent of the household income distribution. The increase in the weighted skills index is substantially larger for poor students; between 45 and 70 percent higher than for richer students (depending on the percentage of transferable resources). This result indicates that student from less affluent background benefit relatively more from the optimal allocation.

Figure 2.7:
Difference between optimal and baseline allocation
of average skills, classroom effects and time investments



Note: The top, middle and bottom panels presents the difference between the optimal and baseline allocation of the average skills, classroom effects and time investments, respectively. See note in Figure 2.6 for additional details.

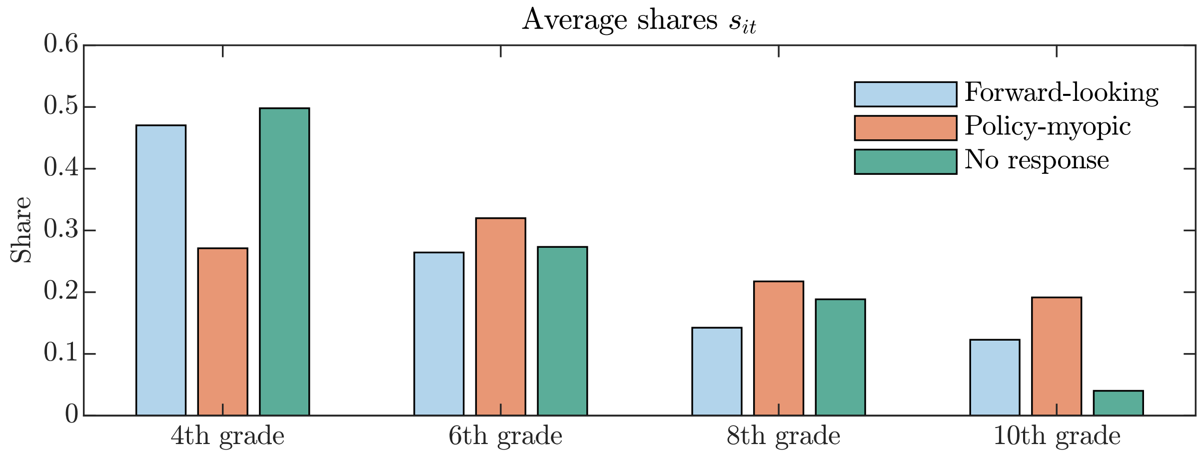
The last exercise aims to quantify the contribution of the behavioral response in the design of the optimal allocation policy. I calculate the optimal allocation defined by equation (2.3.10) for three households' types: 1) forward-looking; 2) policy-myopic; and 3) no response. Figure 2.8 shows the average optimal shares under the three scenarios. In all cases, on average, it is optimal to allocate relatively more resources in lower grades, especially in fourth grade for forward-looking and non-responsive types. The pattern is more pronounced if households do not alter their behavior. However, policy myopic households assign at fourth grade an average share of around 25 percent—almost half of the other two types. Part of the reason is that policy myopic households decrease parental time as a response to additional resources, diminishing the benefits of allocating resources at this grade.[13] Forward-looking households, on the other hand, increase their parental time (on average) when additional resources are allocated in fourth grade. The difference is because forward-looking households internalize that these additional resources imply lower resources in subsequent grades.

I perform a thought experiment to quantify the relevance of households' responses in the design of the optimal allocation. I can characterize the optimal allocation under the assumption that households do not respond. I then calculate the impact of this allocation on child development in the scenario that parents and children do actually respond. Finally, I compare the outcome with the impact of the optimal allocation that internalizes the behavior of households. The differences between the outcomes indicate the importance of the behavioral response in the characterization of the optimal allocation.

In Figure 2.9, I plot the average difference in the weighted skills index between the optimal and baseline allocations. Each bar corresponds to a particular household type (first line of the bar's label) and implements the optimal allocation calculated for an specific household type (second line of the bar's label). For example, the first bar from the left (light blue), shows the average difference between the optimal and baseline allocation

[13] Figure S.2.15 presents the difference of the average classroom effects between the optimal and baseline allocation for each household type.

Figure 2.8:
Average shares of optimal allocation
by household behavioral type

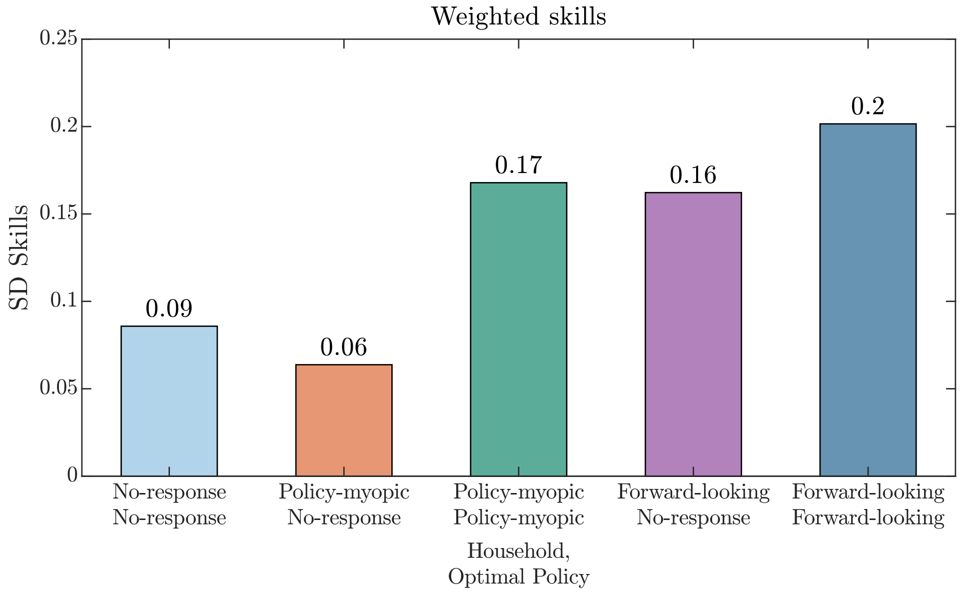


Note: The figure presents the average shares of total transferable resources of the optimal allocation for each household type. No response: households do not respond to different allocations; Policy myopic: households respond as if resources in subsequent grades are the given by baseline allocation; Forward-looking: households are forward-looking and make their decisions understanding the dynamic implications of different allocations. For non-responsive households I simulate skills dynamics using the skill technology (Table S.1.5). For policy myopic household I simulate their choices and skills with the time investment functions (Table S.1.8) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see Section 2.2.1 or the note in Figure 2.6 for additional details).

assuming households do not respond and the allocation is optimal for this household type. Under this scenario, the weighted skills index improves, on average, by 0.09 SD. This allocation would be the one we would characterize as optimal if we did not consider behavioral responses of households at all.

Now, if households actually do respond and this allocation is implemented, the average change in the index would be that of the second (orange) and fourth (purple) bars from the left of Figure 2.9—i.e., 0.06 and 0.16 SD for policy-myopic and forward-looking types, respectively. Meanwhile, if the allocation implemented is optimal for the "correct" household type, the average impact of the optimal allocation equals 0.17 and 0.20 SD for policy-myopic and forward-looking types, respectively, as shown by the third (green) and last (dark blue) bars from the left of of Figure 2.9. Not considering behavior in the characterization of the optimal allocation of school resources implies considerably smaller average improvements

Figure 2.9:
Average differences in weighted skills between optimal and baseline allocation



Note: The figure presents the difference of the average weighted skills between the optimal and baseline allocation. The results are by household type (first line in bar's label) and by implementing the optimal allocation for a particular household type (second line in bar's label). Household types: No response: households do not respond to different allocations; Policy myopic: households respond as if resources in subsequent grades are the given by baseline allocation; Forward-looking: households are forward-looking and make their decisions understanding the dynamic implications of different allocations. For non-responsive households I simulate skills dynamics using the skill technology (Table S.1.5). For policy myopic households I simulate their choices and skills with the time investment functions (Table S.1.8) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 2.2.1 or note in Figure 2.8 for additional details.).

in the weighed skills index, between 25 and 65 percent lower depending on the type of household behavior.[14]

2.4 Conclusions

Motivated by the heterogeneous responses by grade found in chapter 1, in this chapter, I analyze optimal resource allocation policies across education levels to understand how the differential impact by grade of school resources and home investments can be used to

[14] Figure S.2.16 shows the average difference between the optimal and baseline allocations in skills at each grade associated with the scenarios presented in Figure 2.9.

design the optimal allocation of the school resources across grades.

I build and estimate a child development model using an indirect inference approach. The estimates help us understand the extent of the behavior's contribution in the design of the optimal allocation of school resource across grades. I characterize the optimal allocation across grades that maximize weighted skills. Where the weights are given by a regression of college attendance of skills. The optimal allocation improves the (weighted) skills outcome by 0.20 SD with respect to the baseline allocation. On average, it is optimal for schools to invest relatively more in the lower grades. Considering household behavioral responses plays a key role in the design of policies. The characterization of the optimal allocation under the assumption of no behavioral response implies substantially smaller improvements if households do respond to school inputs; between 20 and 65 percent lower depending on the assumptions made regarding the household's expectations process.

Chapter 3

On-the-Job Training and Employer Asymmetric Learning

Chapter summary

In the third chapter, I develop an empirical test for employer asymmetric learning about the productivity gains of On-the-Job (OTJ) training programs. I developed a model of OTJ training and employer learning. I solve the model under two types of learning: (i) asymmetric, where the current employer learns faster than potential employers and (ii) symmetric, where the whole market learns simultaneously. The solution suggests different wage profile predictions under each form of learning. I build a test based on these predictions and implement it on the Chilean Social Protection Survey by estimating a wage equation with interactions of training variables and tenure on the job. The results provide evidence of employer asymmetric learning.

3.1 Introduction

Under the classical microeconomic model of labor supply, firms have the same information about workers' productivity and there is no uncertainty (whether a worker has been employed at a particular firm or not). Several papers have theorized scenarios where the current employer of a worker has better knowledge about her productivity than other potential employers. In the literature, this phenomenon is referred as employer asymmetric learning. The consequences usually involve inefficiencies in several dimensions, such as wages, job mobility and assignment, and human capital accumulation, among others.[1] In this chapter, I build an empirical test for employer asymmetric learning regarding the productivity gain from on-the-job training programs.

Using different strategies a large set of studies have attempted to test the hypothesis that a worker's current employer has better information or learns faster about her productivity (e.g., Acemoglu and Pischke, 1998; Schönberg, 2007; Kahn, 2013; Zhang, 2007; Pinkston, 2009, among others). These papers focus on information asymmetries regarding the innate ability of the worker. If in some degree the information is transmitted between employers, inefficiencies should decline as the worker spends more time in the labor market. The seminal work of Farber and Gibbons (1996) and Altonji and Pierret (2001) evaluates the hypothesis of the employer learning. Furthermore, the work of Schönberg (2007) builds on their methodology to test asymmetric learning regarding workers innate ability and her results suggest that learning might have a non-trivial symmetric component. However, workers' productivity changes along their lives—e.g., via human capital accumulation due to training or learning by doing. For instance, if a worker participates in a training program and her firm has a better understanding of the training's productivity gain than other potential employers, even if every potential employer know the worker's innate ability, inefficiencies might still arise as a consequence of these changes in productivity.

[1] Some examples are Waldman (1984), Greenwald (1986), Waldman (1990), Gibbons and Katz (1991) and Golan (2005).

In this chapter, I focus on employer asymmetric learning about productivity changes due to on-the-job training (OJT). I develop a model of on-the-job training and employer learning. I solve the model under two distinct forms of employer learning: asymmetric and symmetric. In the former, the current firm of the worker—i.e., the firm where the worker received the training—learns the productivity gain from training faster than other potential employers. In the symmetric case, the whole market learns the returns of the training undertaken by workers simultaneously.[2] The model implies different predictions under the two types of learning. These results provide testable implications regarding the way information is transmitted across employers. I develop a test for the presence of asymmetric learning using these predictions. I implement the test on the Chilean Social Protection Survey (SPS) data by estimating a wage equation with interactions of training variables and tenure on the job.

The model consists of a competitive spot labor market, where each period firms make wage offers to workers. Current and potential employers can make wage offers to every worker. Firms and workers can only commit to one-period non-contingent contracts. Training delivers heterogeneous productivity returns and employers learn the return to training after workers participate in the production process. I assume a mass one of workers and infinite number of firms; both types of agents live forever in a discrete time setting. In the initial period workers receive wage offers and some participate in training programs depending on training cost and expected benefits. At the end of the period production takes place and firms learn workers' productivity. Current employers can lay off workers using that information and with some probability workers are separated exogenously. Workers are laid off only if there is asymmetric learning, since outside offers are based on expected productivity. The current firm is not willing to match offers of workers with productivity below the expected productivity and lay offs workers for which it is not possible to make

[2] In the model, I further assume that learning is instantaneous. The difference between asymmetric and symmetric learning is that in the former the current firm learns immediately and potential employers only learn if they hire the worker. In contrast, in the symmetric learning case the entire market learns the workers' productivity right after the training takes place.

profits. Instead, under symmetric learning, the entire marker knows the training returns and firms make offers accordingly. The incumbent firm matches outside offers making workers indifferent. In the symmetric case, workers change jobs only if they are exogenously separated.

The offers to the pool of separated workers also depends on the type of learning. If asymmetric, outside firms do not know which are laid off and which ones exogenously separated, and the framework gives rise to adverse selection. As a result wage offers in the secondhand market are lower than the ones to those who remain at the firm that trained them. However, in the following period, new employers learn their newly hired workers' productivity and makes new wage offers and layoff decisions accordingly. Those who are not laid off will experience an increase in their wages, since the market gathers new information about their productivity. In contrast, in the symmetric case every firm knows every workers' productivity after the first period and make offers accordingly, regardless of whether workers switch jobs or not. These are the key predictions of the model: On the one hand, workers who stay with the firm at which they received training will get wage returns from training and it will be constant across tenure. On the other hand, workers who change jobs lose their wage gains from training due to the adverse selection problem, but if they remain at their current job they recover those wage gains.

To develop the testing strategy, I build on previous work that tests for employer asymmetric learning regarding workers' innate ability. In particular, Schönberg (2007) builds on the employer learning framework of Farber and Gibbons (1996) and Altonji and Pierret (2001). The latter studies suggest that employer learning could be tested by comparing the evolution across workers experience in the labor market of the wage effect of the variables observed and not observed by employers, such as education and ability, respectively. If employers learn to some degree, then for low experience workers, characteristics easily observed by employers, such as education, should be highly priced in the marker. In contrast, characteristics difficult to observe, such as ability, should not affect wages. As experience in the labor market increases the returns to education should get weaker as the returns of ability gets larger.

Schönberg (2007) suggests that in the presence of asymmetric learning the evolution of these variables' effects on wages should be associated with tenure and not experience. Her results suggest that employers learn about their employees' innate ability symmetrically.[3] In a similar spirit, I estimate a wage equation with interacting indicator variables of training with current and previous employer and tenure on the job. Under the predictions of symmetric learning the coefficients of the interaction variables should be equal to zero. In contrast, under the asymmetric setting, the interaction between training at the previous jobs and tenure should be positive.

I build a secondary test following a different strategy. Some studies use exogeneous variation that allows them to identify asymmetric information between current and potential employers. If firms can be discretionary about which workers to lay off, they would first let go those with the lowest productivity. If outside firms do not observe these workers' productivity, this implies adverse selection in the pool of laid off workers. Gibbons and Katz (1991) argue that if this is the case, workers who were laid off due to plant closings should have, on average, higher wages at their new firms, relative to those laid off for other reasons—since these workers are less affected by the adverse selection problem. Comparing these workers groups they find evidence of asymmetric information.

Acemoglu and Pischke (1998) use a similar strategy. They suggest that workers who are drafted in the military should not be subject to the adverse selection problem and their wages should be, on average, larger than those who change jobs for different reasons. Using a sample of young workers from Germany they find evidence of asymmetric information. I perform a test in the spirit of the Gibbons and Katz (1991) strategy. The adverse selection problem should affect trained workers as well; the laid off workers should be those with

[3] It should be noted that under similar but extended frameworks Zhang (2007) and Pinkston (2009) find evidence that employer learning about workers innate ability is strongly asymmetric.

lower training productivity returns.[4] I can then compare the returns of training of workers who were laid off because of plant closing and those separated for other reasons. In the presence of asymmetric information, the former group should have, on average, larger wage returns to training.

The data in the chapter comes from the Social Protection Survey (SPS). The SPS is a nationally representative Chilean household survey with a panel scheme with five waves—2002, 2004, 2006, 2009 and 2015. It contains detailed employment and training histories as well as other socioeconomic characteristics of the respondents. Thus, following Loewenstein and Spletzer (1998) it is possible to estimate the returns to training. Moreover, under specifications that nest the predictions of the model under both types of learning, I can build the empirical test of asymmetric learning. In addition, it is possible to use a similar strategy as Gibbons and Katz (1991) since there is information on reason for job separation. Under the assumption that displacement due to plant closing implies no adverse selection, the return to training that took place at previous jobs should be larger for those displaced because of plant closings. Both sets of results suggest that learning about training returns is asymmetric.

The structure of the chapter is as follows: The next section develops the theoretical framework. It includes the model and the equilibrium under two extremes forms of learning, that is, with employer symmetric and asymmetric learning. Section 3.4 presents the data and descriptive statistics. I present the empirical testing strategy for testing employer asymmetric learning regarding the productivity gain from OTJ programs in Section 3.5. Section 3.6 presents the results. Finally, Section 3.7 concludes the chapter.

[4] In the model there is exogenous separation of workers, however, I assume that potential employers can not distinguish between those workers and the laid off ones. This strategy is in turn somewhat conflicting with the setting of the model. Nevertheless, it should be thought as a particular case of separation shock where firms are aware of this separation. Instead the exogenous separation in the model should be though as a idiosyncratic shocks instead.

3.2 Model of on-the-job training and employer learning

I develop a model of on-the-job training with heterogeneous productivity returns and employer learning regarding the productivity gains. The aim is to obtain wage profile predictions under two types of learning: symmetric and asymmetric. On the one hand, symmetric learning across current and potential employers means that, once an employer learns a worker's productivity following participation in a training program, the whole market does as well. On the other hand, asymmetric learning refers to when the current employer learns faster than outside firms. In particular, I consider the extreme case were the learning of the incumbent firm represents perfect information while outside firms do not learn at all.[5] The different predictions that follow from the two scenarios allow me to empirically evaluate the existence of employer asymmetric learning.

The equilibrium sheds light on the mechanisms by which learning affects training decisions and wage profiles depending on whether all firms learn at the same time or asymmetrically. This section proceeds as follows: First, I outline the main features of the model. Second, I present a characterization of agents and technology, information structure, timing of events, agents' behavior, and equilibrium conditions. Finally, I provide a parametrization and the solution of the model under both learning settings.

3.2.1 Description of the model

The model consists of firms and workers who maximize discounted profits and earnings, respectively. While firms are homogeneous, workers are heterogeneous across innate ability.

At the initial period of the model there is the possibility of a training program which increases the productivity of workers. The productivity gains are heterogeneous, and an

[5]Learning may take different forms, and information can spread across the market in many ways. For the current chapter, I only consider the distinction between the learning of the incumbent firm and outside firms. Furthermore, learning could be gradual. However, the extreme cases I consider make the model more tractable, and the qualitative theoretical results are not affected.

employer learns the change of the worker's productivity only after she participates in the production process at that firm. The assumption about the symmetry of learning plays a role at this point. If the current firm of the worker learns her productivity gain at the same time that any other potential employer, it would be considered symmetric learning. In contrast, if only the current employer learns the return, but it is still unknown to any other potential employer, it is asymmetric learning.

Firms compete for workers and free entry implies ex-ante zero profits. However, it is assumed that firms incur costs to hire workers. Therefore, firms have ex-post monopsony power and pay workers below their productivity, regardless of the worker's training status.

If workers are trained, the presence of asymmetric information and hiring costs allow firms to extract additional rents. For example, if outside firms do not observe the productivity of trained workers, incumbent firms offer wages below productivity and extract additional rents, while workers from whom firms cannot extract rents at equilibrium are lay off. In addition, the model includes exogenous separation; that is, even workers the incumbent firm prefer to retain are separated from the job with some probability. Outside firms know the probability distribution of that event but do not observe if a particular worker was lay off due to her productivity or for exogenous reasons.

In the following subsections I describe the features of the model. In particular, I will focus on the setting with asymmetric employer learning. First, I describe the agents and environment, following with the information structure and timing of the model. Then, I detail the behavior of the agents and the conditions that characterize the equilibrium. Finally, I describe the parametrization and solution of the model under the two learning scenarios.

3.2.2 Agents and environment

There is a mass 1 of workers, heterogeneous across innate ability, denoted by μ and drawn from the distribution Γ. Workers live forever and discount future values with the parameter

β. There is a training cost, c, drawn from the distribution G. Firms offer wages and workers select into training by observing their own cost. It would be not profitable for the worker to get training if she has a sufficiently large training cost.

Firms are homogeneous and there is free entry. The firm where the worker is currently employed will be labeled the *incumbent firm*, while any potential employer will be called an *outside firm*. The terms firm and employer will be used interchangeably. The production function has constant returns to scale, and the output generated by each worker equals their productivity. To hire a worker firms will have to pay a fixed cost denoted by h. Hiring costs play a key role as in Gibbons and Katz (1991) and Hu and Taber (2011). These costs could be interpreted as mobility costs incurred by the worker, hiring costs incurred by the new employer or firing costs incurred by the current employer.[6] Anything that implies an advantage to the current employer with respect to potential outside employers will lead to the same results.

The model includes investment decisions in on-the-job training. Those who are trained increase their productivity by α, which is drawn from a distribution F with density f. The α will also be referred to as the "type" of a worker. Thus, given the realization of c and the expected returns to training, there is a threshold on the support of the training cost distribution G such that every worker with a c below it will be trained. If workers participate in a training program their productivity is $\mu + \alpha$; otherwise it is given by μ.

3.2.3 Information structure

Every agent in the model observes the innate ability of workers, μ, and their associated training cost, c. However, employers only learn α after the production process takes place. It is assumed that α is independent of c and μ. The qualitative results hold without this restriction and it simplifies the analysis. Firms know all labor histories of workers—i.e.,

[6] This assumption seems quite reasonable, especially in a context like Chile. At least in the formal sector, the firing costs in Chile are high. For every year the employee worked at the firm, the employer must give a severance payment of one monthly salary. The payments start running after half a year of contract, and the maximum corresponds to 11 years.

wages and jobs they previously had. Moreover, since employers observe c they also know which workers were and were not trained. Firms know the distribution of training returns, F.

Given the existence of hiring costs, the incumbent firm has ex-post monopsony power and is able to generate positive profits from some workers, while laying off non-profitable workers. Even though outside firms do not observe each worker's α, they know F and the threshold incumbent firms use to make lay off decisions. It is assumed that there is a potential shock by which some employees are separated from the job exogenously. With probability ξ—known by workers and firms—a retained worker is separated from her job due to reasons exogenous to the model. Outside firms, however, cannot distinguish between laid off workers and those who were hit by this shock. This assumption introduces turnover that is not a consequence of asymmetric information. To simplify calculations, and without loss of generality, it is assumed that this shock happens only at the end of the period in which training takes place.

In summary, firms and workers observe everything in the model but the productivity returns from training. Under asymmetric learning, only the incumbent firm learns its trained workers' α after they finish the production process. However, in the symmetric learning setting, once one firms learns the workers productivity, the whole market has that information. By assumption, workers do not know their own α—workers have the same information as outside firms. If this were not the case, outside firms would be able to design wage contracts such that workers would reveal their α. This assumption shuts down this possibility.

3.2.4 Timing

Similar to Hu and Taber (2011), workers live forever but all the interesting features of the model happen in the first three periods. In a finite period version of the model, wage profiles will be affected by the retirement decision. The assumptions imply that the training decisions are made far before retirement. The events that occur at each period are as follows:

- *Period 1:* First, each worker gets a realization of μ and c drawn from the distributions Γ and G, respectively. Those with training costs below a certain threshold are trained. Second, firms make offers to workers, competing away any discounted value of expected future profits and workers choose which one to accept. Third, workers participate in the production process and employers learn their trained workers' productivity. Finally, employers make lay off decisions and there is a potential separation shock to retained workers.

- *Period 2:* At the beginning of the period firms make offers and workers decide which one to accept. The production process takes place and firms—if they did not know them already—learn their workers' productivity and lay off the ones who are not profitable (there is no separation shock at this period or beyond).

- *Period 3 and all later periods:* Firms make offers and workers decide which one to accept. Workers participate in the production process, and there are no more layoffs.

It is assumed that a firm only lays off workers at the end of the period in which it learned their productivity. It may be profitable for firms to lay off workers in later periods; however, they should be few in number and the assumption of no later layoffs simplifies the calculations. It is also assumed for similar reasons that the separation shock arrives only at period 2.

3.2.5 Behavior of workers and firms

Workers are risk neutral and they maximize the expected present value of their earnings. By assumption trained workers do not know their own realization of α. Workers have the same knowledge than outside firms about their productivity. Therefore, workers behavior is quite simple: it only involves choosing the optimal offer at each period.

Risk neutral firms maximize discounted expected profits and employ different groups of workers; (i) those who were not trained, (ii) those who were trained, and the firm knows their productivity and (iii) those who were trained, and the employer does not know their productivity. Constant returns to scale imply that firms could have any combination and number from these groups of workers—i.e., there could be one worker at each firm, a single firm hiring every worker, or any combination in between.

When firms make offer to workers at the beginning of period 1 they know which workers will receive training but they do not know the realizations of α. Firms will compete away any positive expected profit. After trained workers participate in the production process, their employers learn their productivity. Since outside firms do know this information in the asymmetric learning case and there are hiring costs, incumbent firms lay off workers from whom they cannot make profits. However, given the possibility of the separation shock, workers who leave the firm will be a combination of laid off workers and those who received the shock. Outside firms are not be able to distinguish between them but will use this information.

At period 2, outside firms make offers knowing which workers are laid off or separated and which ones are retained. There are workers from all three groups in the economy. First, workers produce, and firms learn workers' productivities. Second, at this period there are two kinds of incumbent firms: those which already made their lay off decisions and those which have not done so yet. The latter will decide which worker lay off, retaining only those who are profitable.

By assumption there are no more layoffs after period 2. Thus, at period 3 and in all later periods, firms make wage offers and workers decide which offer to accept maximizing their expected present value of earnings.

3.2.6 Equilibrium

The equilibrium in the model comes from four conditions:

1. *Outside firms' wage offers are determined such that firms earn on expectation zero profit.*
2. *Incumbent firms' wage offers are chosen such that workers are indifferent between staying or leaving.*
3. *Firms retain workers for whom it is profitable to do so.*
4. *Workers make choices such that they maximize the expected present discounted value of earnings.*

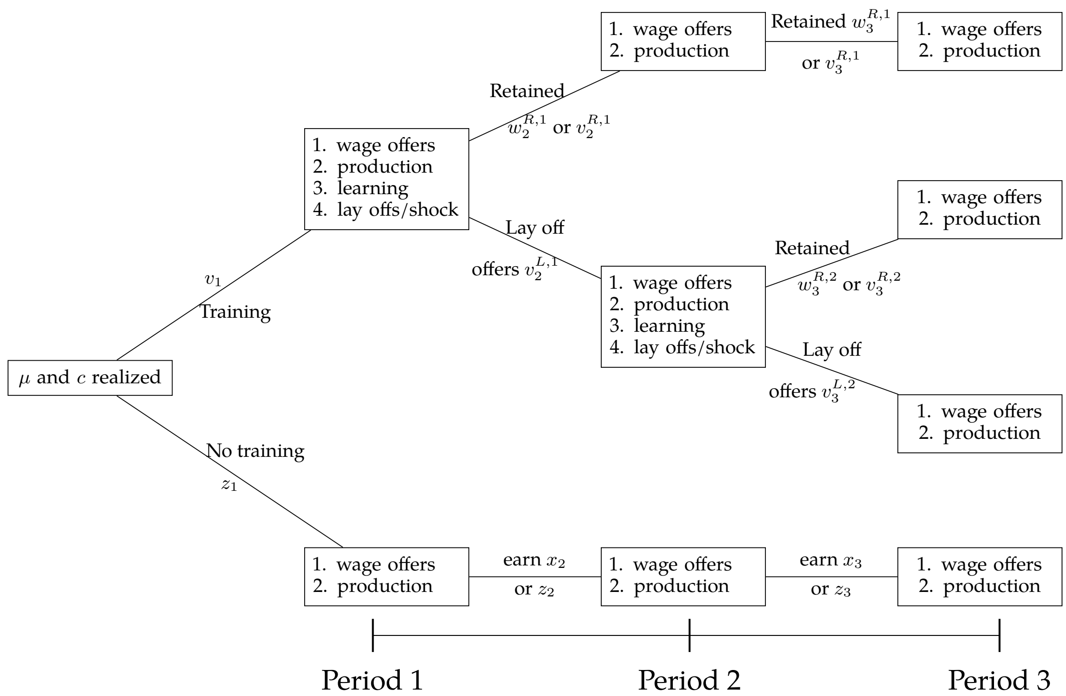
3.2.7 Parametrization and solution

Let $F \sim \text{Uniform}(0, \overline{\alpha})$ and $G \sim \text{Uniform}(0, \overline{c})$. The functional forms of these distributions has implications for the model; however, it makes the calculations simpler, and the results hold for any well-behaved distributions. It is assumed that hiring costs are small enough such that it is profitable to hire lay off workers, but large enough that is not profitable to lay off workers at the third period. To solve the model, it is not necessary to parameterize the distribution of workers' innate ability; it will just be denoted by Γ.

Figure 3.1 presents the timeline of the first three periods of the model—all further periods will be exactly the same as the third one. Some notation clarification: letters w and v denote incumbent and outside firms' offers to trained workers; letters x and z refer to incumbent and outside firm offers to not trained workers. First period wage offers to (future) trained and untrained workers are denoted by v_1 and z_1, respectively. After period 1, workers who did not receive training are offered x_t by their current firm and their outside offer is denoted by z_t, where $t \geq 2$ index periods.

Wage offers to trained workers require additional notational detail. Let $w_t^{k,j}$ denote a wage offer of an incumbent firm at period t, where the supra-index k denotes the separation

Figure 3.1: Timeline of the model



status—$k = R$ and $k = L$ for retained and laid off (or exogeneously separated) workers, respectively— and $j \geq 1$ indexes the period at which the worker was retained for the first time or, if $k = L$, it states the period the worker was laid off.[7] Outside firms observe the layoff/retained status of the worker, thus, they will offer different wages to each group. In summary, there are three possible offers associated with trained workers; Incumbent firms' offer, $w_t^{R,j}$, outside offers to retained workers, $v_t^{R,j}$, and outside offer to laid off workers, $v_t^{L,j}$.

Under the previous notation and parametrization, I solve the model in the following subsection. First, I describe the solutions in an environment with asymmetric employer learning, and second, the solution under symmetric learning.

[7]Incumbent firms' wage offers to retained workers will vary depending on the period they were retained for the first time. Firms will track this information because it implicitly assigns the worker to a particular group of types.

Asymmetric learning equilibrium

This section presents the solution of the model under asymmetric learning. I solve the model by backward induction. First, I develop the solution of the non-trained workers' problem.

The productivity of these workers is known. When a firm hires a worker it incurs the costs denoted by h. Nevertheless, once the worker is hired, the firm has monopsony power—any potential employer will have to pay the hiring cost. However, any positive profits the firm may get are competed away in the first place due to the free entry assumption. Thus, at any period t non-trained workers will be offered their productivity minus the hiring cost plus the present value of expected profits of future periods.[8] Formally, the outside wage offers at period t are:

$$z_t = \mu - h + \beta \mathrm{E}_{\xi}(\pi^N_{t+1}), \tag{3.2.1}$$

where π^N_{t+1} denotes firm' discounted value of profits from period $t+1$ and on and the expectation operator is over the distribution of the exogenous separation shock, ξ. I define $\pi(\alpha, y_t, y_{t+1}, \ldots)$ as the profits discounted value from period t and beyond a firm obtains if it retains a worker of type α forever with the wage sequence $(y_t, y_{t+1}, \ldots)$. These profits do not include hiring costs. To reduce notation, I denote the wage sequence $(y_t, y_{t+1}, \ldots)$ as $y(t)$. Formally:

$$\pi(\alpha, y(t)) = \sum_{\tau=0}^{\infty} \beta^{\tau}(\mu + \alpha - y_{t+\tau}). \tag{3.2.2}$$

Thus, $\pi^N_{t+1} \equiv \pi(0, z(t+1))$. The current employer has no incentives to offer more than a tiny bit above the outside offer and it faces the same problem at every period. Hence, the solution at every t is the same and for every $t \geq 2$,

$$z_t \;=\; \mu - (1-\beta)h, \tag{3.2.3}$$

[8] In equilibrium, the only uncertainty about workers who were not trained comes from the exogenous separation shock that hits at the end of period 1; there is no uncertainty after μ and c are realized.

and $\pi(0, z(t+1)) = h$. Incumbent and outside firms' wage offers to untrained workers will be $z_t = \mu - (1-\beta)h$ for every $t \geq 2$. In contrast, given the exogenous shock probability ξ at the beginning of period 2, for $t = 1$ we have that $z_1 = \mu - [1 - \beta(1-\xi)]h$.

The wage determination of trained workers is more interesting. To solve for the firms' offers I follow a similar path as for non-trained workers: first, the retained workers' outside offers, and so the retention offer by the incumbent firms; second, lay off decisions; and third, outside offers to separated workers. Consider a firm that at period 2 wants to poach a worker from a firm that retained her at period 1. The expected value of a worker is the sum of her expected productivity at each period minus the hiring cost, h,

$$\sum_{t=2}^{\infty} \beta^{t-2} \Big[\mu + \mathrm{E}(\alpha \mid \alpha \geq \alpha^*)\Big] - h. \tag{3.2.4}$$

where α^* is the cut-off level in the support of F where every trained worker below it is laid off in period 1 and those above it are retained. This equation is crucial for the determination of the incumbent firms' wage offers. Outside firms will compete away any expected future profit to attract the workers. Thus, the outside offer, $v_2^{R,1}$, is given by the worker' expected value. The current employer is not willing to offer more than a tiny bit above it—i.e., at equilibrium $w_t^{R,1} = v_t^{R,1}$ for any $t \geq 2$.

If an outside firm poaches a worker with productivity $\mu + \alpha$ at period 2, the period 3 discounted value of equilibrium profits is $\pi(\alpha, w^{R,1}(3))$.[9] It is worth noticing that, if the firm poaches the worker, in the following periods it will pay exactly the same (equilibrium) wages that the previous employer would have paid, $w_t^{R,1}$ for $t \geq 3$. Thus, the outside offers to a retained worker at period 2, $v_2^{R,1}$, are:

$$v_2^{R,1} = \mu + \mathrm{E}(\alpha \mid \alpha \geq \alpha^*) - h + \beta \mathrm{E}(\pi(\alpha, w^{R,1}(3)) \mid \alpha \geq \alpha^*), \tag{3.2.5}$$

[9] In equilibrium this does not happen. Workers will not accept outside offers if they are not laid off; however, outside offers are crucial for the determination of their wages in equilibrium.

and the incumbent firm will match the offer—i.e., $w_2^{R,1} = v_2^{R,1}$.[10] Furthermore, it chooses α^* such that it is indifferent to retaining the worker or not. That is,

$$\mu + \alpha^* - w_2^{R,1} + \beta\pi(\alpha^*, w^{R,1}(3)) = 0. \tag{3.2.6}$$

It should be noted that profits in the previous equation are certain, there is nothing that the current employer does not know about the worker. Those who are laid off receive offers from outside firms that only know that their α is below α^*. Thus, these offers are expressed as

$$\begin{aligned} v_2^{L,1} \quad = \quad & \mu - h + \frac{F(\alpha^*)}{F(\alpha^*) + [1 - F(\alpha^*)]\xi} \times \\ & \left[\mathrm{E}(\alpha \mid \alpha < \alpha^*) + \beta\frac{[F(\alpha^*) - F(\alpha^{**})]}{F(\alpha^*)}\mathrm{E}(\pi(\alpha, w^{R,2}(3)) \mid \alpha^{**} \leq \alpha < \alpha^*)\right] \\ & + \frac{[1 - F(\alpha^*)]\xi}{F(\alpha^*) + [1 - F(\alpha^*)]\xi} \times \left[\mathrm{E}(\alpha \mid \alpha \geq \alpha^*) + \beta\mathrm{E}(\pi(\alpha, w^{R,2}(3)) \mid \alpha \geq \alpha^*)\right] \end{aligned} \tag{3.2.7}$$

where α^{**} is the cut-off level by which the new firm makes its lay off decisions—i.e., it will lay off all workers with α below α^{**}, and retain them otherwise. The share of trained workers that change jobs at the end of period 1 is $F(\alpha^*) + [1 - F(\alpha^*)]\xi$. Among those, a fraction $F(\alpha^*)/(F(\alpha^*) + [1 - F(\alpha^*)]\xi)$ was laid off and the complement $[1 - F(\alpha^*)]\xi/(F(\alpha^*) + [1 - F(\alpha^*)]\xi)$ was exogenously separated employees.

Furthermore, $[F(\alpha^*) - F(\alpha^{**})]/F(\alpha^*)$ is the fraction of laid off workers who will not be laid off at the end of period 2. These workers at period 3 have an outside offer given by its expected productivity minus hiring costs plus any positive expected discounted value of

[10]The offer $v_2^{R,1}$ does not include the probability of exogenous separation in the expectation of profits since that kind of separation only happens at the end of period 1 by assumption.

future profits, formally:

$$
\begin{aligned}
v_3^{R,2} \quad = \quad & \mu - h + \frac{F(\alpha^*) - F(\alpha^{**})}{F(\alpha^*) - F(\alpha^{**}) + [1 - F(\alpha^*)]\xi} \times \\
& \qquad \Big[\mathrm{E}(\alpha \mid \alpha^{**} \leq \alpha < \alpha^*) + \beta \mathrm{E}(\pi(\alpha, w^{R,2}(4)) \mid \alpha^{**} \leq \alpha < \alpha^*)\Big] \\
& + \frac{[1 - F(\alpha^*)]\xi}{F(\alpha^*) - F(\alpha^{**}) + [1 - F(\alpha^*)]\xi} \times \Big[\mathrm{E}(\alpha \mid \alpha \geq \alpha^*) + \beta \mathrm{E}(\pi(\alpha, w^{R,2}(4)) \mid \alpha \geq \alpha^*)\Big]
\end{aligned}
\tag{3.2.8}
$$

New incumbent firms match the previous offers, $w_3^{R,2} = v_3^{R,2}$. Thus, firms choose α^{**} such that they are indifferent between retaining and laying off the worker with $\alpha = \alpha^{**}$. That is,

$$
\mu + \delta\alpha^{**} - w_3^{R,2} + \beta\pi(\alpha^{**}, w^{R,2}(4)) = 0. \tag{3.2.9}
$$

Workers with $\alpha < \alpha^{**}$ are laid off at the end of period 2. Using that information outside firms make offers,

$$
v_3^{L,2} \quad = \quad \mu - h + \mathrm{E}(\alpha \mid \alpha < \alpha^{**}) + \beta \mathrm{E}(\pi(\alpha, w^{R,3}(4)) \mid \alpha < \alpha^{**}) \tag{3.2.10}
$$

The last condition needed to close the model is to define period 1 wage offers to those who are trained. Firms know the value of c, so they know which workers will be trained. Nevertheless, potential employers do not know the productivity return of each worker. They offer workers their expected productivity plus the discounted value of expected future profit from period 2 and beyond, minus the hiring and training costs:

$$
v_1 = \mu + \mathrm{E}(\alpha) - h + \beta[1 - F(\alpha^*)](1 - \xi)\mathrm{E}(\pi(\alpha, w^{R,1}(2)) \mid \alpha \geq \alpha^*) - c. \tag{3.2.11}
$$

The assumption that workers are credit constrained sets the training cost threshold's value at which workers are trained or not. Let c^A denote the threshold such that workers with a lower training cost get training and those above it do not, where the superscript A

denotes the presence of asymmetric information. In the presence of symmetric information the threshold will be denoted by c^S.

Thus, with credit constrained workers and employer asymmetric learning, $c^A = \beta[1 - F(\alpha^*)](1-\xi)\mathrm{E}(\pi(\alpha, w^{R,1}(2)) \mid \alpha \geq \alpha^*)$. There are two sources of inefficiency, one comes from the fact that under asymmetric learning firms do not collect profits from all workers who are trained—only from a fraction $[1 - F(\alpha^*)](1-\xi)$. However, as other firms can extract rents from those workers, it would be optimal for them to contribute to additional workers' training, but there is no channel by which they could it. The second source is trivially the financial constraint of the workers. The value of the cut-off is[11]

$$c^A = \beta(1-\xi)\frac{2(1-\beta)h}{\overline{\alpha}}h \tag{3.2.12}$$

Note that if workers are not credit constrained, investment in training is optimal regardless of whether learning is symmetric or asymmetric. In the first period, workers receive wages that include all expected benefits, so perfect financial markets lead to optimal training decisions. However, if workers are financially constrained to some degree, asymmetric learning implies under-investment in human capital. Nevertheless, the predictions of the model under the two types of learning does not differ by whether workers are credit constrained or not; only the fraction of trained workers differs under the two scenarios.

Symmetric learning equilibrium

Symmetric learning implies that after the incumbent firm knows the productivity return of training every other potential employer also knows the exact value of each worker's productivity. The symmetric case is an environment where the labor market is competitive but there are hiring costs. These costs imply ex-post monopsony power, however, in this environment firms cannot make additional profits out of the workers' training returns,

[11] See Appendix E.

unlike the asymmetric learning setting. Thus, there are no lay offs, workers separate only exogenously at the end of the first period and remain at their job from period 2 on in equilibrium.

At period 2, outside offers to untrained workers are the same as in the asymmetric environment. At the beginning of period 2, some trained workers were separated from their job due to exogenous reasons with probability ξ and firms make them offers. In the symmetric case, separation information is irrelevant to outside firms because they observe the worker's productivity and make offers to retained and separated workers accordingly. As in the asymmetric environment they compete away any discounted value of future profits and subtract the hiring cost, so that,

$$v_2 = \mu + \alpha - h + \beta\pi(\alpha, w(3)). \tag{3.2.13}$$

The previous offer will be exactly the same at each period and the current employers match it—i.e., $w_t = v_t$ for every $t \geq 2$. Profits are given by:

$$\pi(\alpha, w(3)) = \sum_{t=3}^{\infty} \beta^{t-3} \left(\mu + \alpha - w_t\right). \tag{3.2.14}$$

Thus, offers made by incumbent firms—and outside firms— are defined as $w_t = \mu + \alpha - (1-\beta)h$ for every $t \geq 2$. At period 1, competition among firms determines the wages (future) trained workers will earn:

$$v_1 = \mu + E(\alpha) - h + \beta(1-\xi)E[\pi(\alpha, w(2))] - c \tag{3.2.15}$$

where the profits' discounted value is given by $\pi(\alpha, w(2)) = h$. If workers are credit constrained, the training costs threshold is given by the profits' discounted value—i.e., $c^S = \beta(1-\xi)h$.

3.3 Predictions of the model

The model predicts different patterns of wages across tenure under employer asymmetric and symmetric learning. Those results can be used to develop a testing strategy for asymmetric learning. In particular, if employers learn the training return symmetrically, the wages will not vary across tenure levels beyond the training period, whether the worker is at the training firm or at a different one. After a firm learns the productivity of a worker, that information spreads through the whole market and competition implies that workers collect the returns to training at any firm.

In the asymmetric environment the previous result for trained workers does not hold anymore. The wage profile with tenure is different if the worker stays at the training firm or if she moves to a new job. If the worker is retained at the training firm, after the training period there is no change in wages, however, after the training process is over and learning takes place wages go from v_1 to $w_2^{R,1}$, that is the following increase

$$w_2^{R,1} - v_1 \quad = \quad \mathrm{E}(\alpha \mid \alpha \geq \alpha^*) - \mathrm{E}(\alpha) + (F(\alpha^*) + [1 - F(\alpha^*)]\xi)\beta h + c \tag{3.3.16}$$

In contrast, those who change jobs have lower wages than in the previous firm. Their wages are lower due to the adverse selection effect. The wage difference is given by $v_2^{L,1} - w_2^{R,1}$. Furthermore, if those workers are retained at their new job, they get a wage increase in the following period $w_3^{R,2} - v_2^{L,1}$, and adverse selection in job separation guarantees that this difference is positive.

3.4 Data

The Social Protection Survey (SPS) is a Chilean nationally representative survey with a panel scheme with five waves— 2002, 2004, 2006, 2009, and 2015. Each wave surveyed around 16,000 individuals who were older than 18 years old (except for 2002 when the threshold was 15 years old). It contains detailed employment and training histories as well as other socioeconomic characteristics of the respondents. In particular, every employment, unemployment, and inactivity spell since each individual was 18 years old (or since their age in 1980 if they were older at that time). Furthermore, the survey reports spell characteristics—type of contract, social security information, average monthly earnings, hours worked, separation reasons and inactivity reasons, among others. Information on wages and hours worked is available starting in 2002.

The SPS includes a section that collects information about formal training undertaken by the respondents throughout their lives and characteristics of that training— e.g., funding source, institution carrying out the program and length. There are some drawbacks; most notably the information of training programs is only available for the three main programs undertaken before the wave 2002 or between waves.[12] Thus, at least to some degree we cannot observe all the training programs undertaken by each individual. However, only 10 percent of the sample reports having more than 3 training programs between interviews, suggesting that the issue may not bias the results to a large extent.

Since the focus is on employer-sponsored training, the sample includes only wage and salary workers aged between 18 and 65 years old. Furthermore, I drop respondents without information when they were younger than 20 years old; to avoid individuals lacking complete labor or training histories. In addition, I do not include observations on the top 1 percent of wages or those whose wages change between to jobs belongs to the top 1 percent of the wage change distribution. Lastly, I drop workers with weekly hours worked above 90 (top 1 percent) or below 9 (bottom 1 percent) as well. This restriction is made since some wages and wage growth seem large and are likely measured with error, as in the case of hours worked during a week.

Descriptive statistics on workers' characteristics appear in Table 3.1, where column (3)

[12] Each wave reports detailed information on the three most relevant training programs in the time interval between surveys (in 2002 since the respondent was 18 years old or since 1980 if older at the time). Nevertheless, there is information of whether there was participation in 4 or more programs.

shows the statistics for the full sample and columns (1) and (2) split the sample in two: individuals with (Column 1) and without (Column 2) employer-sponsored training. In this chapter the term "training" refers only to employer-sponsored training. The first row (Percent) shows the proportion of trained workers, 25 percent of the sample has gone through a training program. The average age at the time of the 2004 wave was 32 years old, and on average trained workers are 2 years older and a larger share are men.

These two groups have marked differences regarding their labor market outcomes. As expected, trained workers, on average, earn a higher hourly wage while weekly hours worked are similar (but have a higher variance for those with no OJT). Untrained individuals have a larger share of part-time and informal labor. While informal labor accounts for 5 percent of OJT workers, it is 18 percent for untrained ones. More than a quarter of trained employees are in a union while about 12 percent of their untrained counterparts are.

There is a gap of labor experience of almost 5 years in favor of trained workers. Job duration is another dimension where these groups differ. Trained employees remain at a job, on average, 3.5 years more (3.7 and 6.3 years for untrained and trained workers, respectively). Parent (1999) finds similar results for United States National Longitudinal Survey of Youth (NLSY) 1979 sample, however, that is a younger sample and thus levels are lower. Since average age does not differ much between groups, this may be explained by shorter inactivity or unemployment spells of OJT workers. Moreover, the average number of previous jobs held do not present substantial difference.

The model suggests different patterns of wages across tenure if workers are at the firm where they were trained or at a different one. Table 3.2 has information for workers without training, trained at the current firm or at a previous one.[13] Wages are higher for workers who are currently at the firm at which they participated in the training program, and having training, on average, implies a higher hourly wage. On average, tenure is substantially

[13] Table 3.2 presents two subgroups of trained workers at previous firms. Those who received training at a previous firm and may or may not have participated in training programs at the current firm, and another groups of workers who only got training at a previous firm.

Table 3.1:
Characteristics of workers by On-the-Job training status

	Without OJT (1)	With OJT (2)	Total (3)
Percent	74.4	25.6	100.0
	(43.7)	(43.7)	
Age (at 2004)	31.8	33.9	32.2
	(7.4)	(6.5)	(7.3)
Male (percent)	54.0	62.1	56.1
	(49.8)	(48.5)	(49.6)
Log hourly wage (CLP)	8.3	8.7	8.4
	(0.6)	(0.6)	(0.6)
Weekly hours worked	45.7	46.4	45.9
	(10.3)	(8.2)	(9.8)
Tenure (in years)	3.7	6.3	4.4
	(4.8)	(6.0)	(5.2)
Experience (in years)	10.9	14.5	11.9
	(7.9)	(7.2)	(7.9)
No. previous jobs	3.6	3.3	3.5
	(3.9)	(3.0)	(3.7)
Part-time (percent)	9.1	4.2	7.8
	(28.7)	(20.0)	(26.8)
Informal (percent)	18.2	5.0	14.8
	(38.6)	(21.8)	(35.5)
Union (percent)	12.1	27.7	16.1
	(32.6)	(44.7)	(36.7)
No. of observations	21,055	7,254	28,309
No. of individuals	7,894	2,571	10,465

Source: Own calculations based on microdata from Social Protection Survey (SPS) 2002-2015.
Note: Standard deviations in parentheses. OJT = On-the-Job Training. Wages are denominated in Chilean pesos (ClP). Definitions: Informal if does not make contributions to social security; Part-time if works less than 35 hours a week. Sample includes wage and salary workers aged between 22 and 65 years old. Individuals for whom there is no information before they were 20 years old are not included.

higher for worker at the training firm (around 5 years higher). As expected, the number of jobs is lower for these groups as well. Finally, labor market experience presents similar levels for both groups of trained workers. The lower panel of Table 3.2 presents an *F-test* of the null of no differences in means among the trained workers groups, and all reject the null for every usual test size.

Table 3.2:
Characteristics of workers by current and previous employer on-the-job training

	Log Hourly Wage	Tenure	Experience	No. Previous Jobs	No. of Obs.
TC	8.8	8.6	14.3	2.3	4,505
	(0.6)	(6.2)	(7.3)	(2.3)	
TP	8.7	2.9	15.0	4.7	3,388
	(0.6)	(3.1)	(7.2)	(3.2)	
TP (not TC)	8.6	2.4	14.9	4.9	2,748
	(0.6)	(2.7)	(7.1)	(3.3)	
No training	8.3	3.7	10.9	3.6	21,053
	(0.6)	(4.8)	(7.9)	(3.9)	
All	8.4	4.4	11.9	3.5	28,306
	(0.6)	(5.2)	(7.9)	(3.7)	
F-test TC=TP	16.5	128.0	9.1	72.0	
p-value	0.0000	0.0000	0.0000	0.0000	
F-test TC=TP (not TC)	66.7	3172.2	14.0	995.3	
p-value	0.0000	0.0000	0.0002	0.0000	

Source: Own calculations based on microdata from Social Protection Survey (SPS) 2002-2015.
Note: Standard deviation in parenthesis. TC = On-the-Job Training with current employer. TP = On-the-Job Training with previous employer. Wages are denominated in 2015 Chilean pesos. Experience and Tenure are measured in years. Sample includes salary workers aged between 22 and 65 years old. Not included individuals for whom there is no information before they were 20 years old.
Source: Own calculations based on microdata from Social Protection Survey (SPS) 2002-2015.
Note: Standard deviations in parentheses. OJT = on-the-job training. TC = On-the-Job Training with current employer. TP = On-the-Job Training with previous employer. Sample includes wage and salary workers aged between 18 and 65 years old. Sample does not include individuals for whom there is no information when they were younger than 20 years old.

Table 3.3 shows the distribution of employees across job separation reasons. Columns (1) and (2) present the distribution for untrained and trained employees, respectively, while column (3) accounts for the full sample. Most workers are separated from their jobs because their contract ended, they quit, or because they were fired. The previous statement is true for both sub-samples. However, the share of fired workers is larger among trained workers, while the "contract ending" category shows the opposite relationship. Separation by plant closing is about 3.5 percent of all separation, and this proportion is not different by training status. Column (4) shows the share of workers who received training at each category. Almost 11 percent of workers displaced due to plant closing received training.

In the following section, I describe the empirical strategy used to test for the presence

Table 3.3:
Separation reasons share by On-the-Job training status

	Without OJT	With OJT	Total	Percent OJT
	(1)	(2)	(3)	(4)
Mutual arrangement	6.2	6.9	6.3	21.3
	(24.0)	(25.4)	(24.3)	(41.0)
Quit	20.7	20.0	20.5	18.9
	(40.5)	(40.0)	(40.4)	(39.2)
Contract ended	34.0	23.3	31.9	14.2
	(47.4)	(42.3)	(46.6)	(34.9)
Fired	15.3	25.1	17.2	28.3
	(36.0)	(43.4)	(37.7)	(45.1)
Firm shut down	3.6	4.0	3.7	20.9
	(18.7)	(19.5)	(18.8)	(40.7)
Found better job	5.9	7.0	6.1	22.2
	(23.5)	(25.5)	(23.9)	(41.6)
Health reasons	1.7	1.2	1.6	14.8
	(12.8)	(10.9)	(12.5)	(35.6)
Retired	0.1	0.3	0.1	50.0
	(2.8)	(5.7)	(3.6)	(51.3)
Incidental event	2.8	2.2	2.7	16.3
	(16.4)	(14.8)	(16.1)	(37.0)
Other	9.9	10.0	9.9	19.5
	(29.9)	(30.0)	(29.9)	(39.7)
No. of observations	12,772	3,075	15,847	15,847

Source: Own calculations based on microdata from Social Protection Survey (SPS) 2002-2015.
Note: Standard deviations in parentheses. OJT = On-the-Job Training. Sample includes wage and salary workers aged between 18 and 65 years old. Sample does not include individuals for whom there is no information when they were younger than 20 years old.

of employer asymmetric learning about the productivity returns to training.

3.5 Empirical Implementation

3.5.1 Statistical strategy

In this section I describe how the predictions of the model can be taken into the data. Farber and Gibbons (1996) and Altonji and Pierret (2001) developed an strategy to test employer learning; they argued that the effect on wages of variables correlated with ability not observed by employers, such as the Armed Forces Qualification Test (AFQT) in the NLSY, will increase over time, while the effect of those easily observed by employers—e.g., education— should get weaker over time. Schönberg (2007), Zhang (2007), and Pinkston (2009) extend the analysis to test asymmetric learning by comparing the effects of these variables across tenure and experience levels.

To test for asymmetric learning regarding training's productivity returns, I will use a strategy that follows the same spirit as the previous literature. My model predicts that training carried out at previous jobs should have lower wage effects at the beginning of a new employment spell. However, the effect should increase with tenure at the job.

The empirical implementation follows a richer model than the theoretical model presented in the previous section. The simplicity of the latter aims to focus the attention on the consequences of the types of employer learning. The empirical model, to test the predictions under different types of employer learning, considers a richer setting to control for other factors. First, the empirical implementation considers returns to tenure and experience, while the theoretical model assumes no returns to experience or tenure. Second, it allows for a more flexible relationship between training returns and tenure—in the theoretical model, productivity returns from training are constant across tenure. For example, we could add to the theoretical model a positive trend in the productivity return of training. In this scenario, we draw the predictions under asymmetric learning by comparing the difference in the trend with respect to the trend under symmetric learning. Third, the theoretical model assumes no additional separations after period 2 which simplifies the calculations but does not alter the predictions. That is, if these features are added to the theoretical model, the employer learning predictions under the two types of learning hold.

Thus, consider the following log hourly wage equation for an individual i on a job j at the survey wave t:

$$\log w_{ijt} = TC_{ijt} \times \delta^c + TP_{ij} \times \delta^p + TC_{ij} \times \text{Tenure}_{ijt} \times \delta^{T,c} + TP_{ij} \times \text{Tenure}_{ijt} \times \delta^{T,p} + \mathbf{X}'_{ijt} \times \beta + \mu_i + \varepsilon_{ijt}, \quad (3.5.17)$$

where w_{ijt} is the hourly wage, TC_{ijt} and TP_{ij} are equal to one if the worker received training at current and previous job, respectively, and zero otherwise. Tenure$_{ijt}$ equals tenure of worker i at job j at survey's wave t. $\mathbf{X}_{ijt}$ is a vector of worker and job observable characteristics (including tenure and its squared). Individual heterogeneity is capture by μ_i, and it can be thought as workers' innate ability. Finally, ε_{ijt} is an error term. The coefficients δ^c, δ^p, $\delta^{T,c}$, $\delta^{T,p}$, and β represent the effects of the associated variables on the log hourly wage. There have been studies that relate training return and tenure. Lentz and Roys (2015) develop a setting in which search frictions and firm heterogeneity imply that even if training is specific or general, the workers will collect the return as new job offers arrive, and the result holds at the training firm or future firms. These results would suggest that $\delta^{T,c}, \delta^{T,p} > 0$.

Barron et al. (1989), among others, point out a possible bias of estimating δ^p and δ^c with OLS as a result of correlation of unobserved individual heterogeneity, μ_i, and training variables TC_{ijt} and TP_{ij}—with more able workers more likely be trained.[14]

Following e.g. Loewenstein and Spletzer (1998), I estimate the wage equation including individual fixed effects, which control for μ_i. Thus, if there is employer asymmetric learning, it should be the case that $\delta^{T,p} > 0$ (or $\delta^{T,p} \neq \delta^{T,c}$ if $\delta^{T,c} > 0$. While if this coefficient equals zero, it would be evidence of symmetric learning. In the following section, I present the results of this exercise.

I perform an additional test following the Gibbons and Katz (1991) strategy using

[14] Job match heterogeneity could cause a similar issue (e.g. Loewenstein and Spletzer, 1998; Parent, 1999, 2003). However, the specification does not control for this component.

information on job loss due to plant closings. If workers are separated from their jobs due to plant closing, there should not be adverse selection, and so the effect of training in previous jobs should be larger for workers for this reason. This suggests the following specification:

$$\log w_{ijt} = TC_{ijt} \times \delta^{c} + TP_{ij} \times \delta^{p} + TP_{ij} \times \text{Plant Closing}_{ij} \times \delta^{PC} + \mathbf{X}'_{ijt} \times \beta + \mu_i + \varepsilon_{ijt}, \quad (3.5.18)$$

where Plant Closing_{ijt} takes the value 1 if the worker was separated from the previous job due to a plant closing, and zero otherwise. Under my assumptions, a positive δ^{PC} suggests the presence of asymmetric information between employers.

3.6 Results

The main testing strategy follows the specification in equation (3.5.17)—i.e., estimating the effects of training in previous jobs on wages across tenure. To get further evidence, I performed an additional test by estimating the specification stated in equation (3.5.18)–i.e., comparing the effects of previous training on wages for workers who were separated due to plant closings against those who were separated for other reasons. The results suggest employers learn about training productivity gains asymmetrically. Firms who train workers seem to learn changes in productivity because of training programs faster than outside firms.

Table 3.4 shows the training effect estimates. The sample consists of salary workers between 18 and 65 years old with information available before they turn 20 years old. Every specification includes controls for observable characteristics.[15] As before, TC and

[15]The observable characteristics are: tenure, tenure squared, experience, experience squared, age at the beginning of the job and its square, part-time job, informal job, number of previous jobs, union status, first job indicator, marital status, region, type of employment (permanent, temporal, fixed time, by task or other), economic activity, occupation, survey year, accumulated time on unemployment, inactive, and in self-employment, participation in emergency employment program indicator, transitions between economic activities and occupations, accumulated time on each economic activity and occupation, separation reasons in previous jobs, attended educational institution for training indicator, and training in current and previous

TP denote training at current and previous jobs. Column (1) shows the estimates without considering individual fixed effects, while column (2) presents the results including individual fixed effects. Training returns drop dramatically after controlling for individual heterogeneity. The coefficient of TC goes from approximately 16 to 5 percent of the hourly wage, while the for TP, it decreases 8 percentage points (pp), and it is not statistically significant after controlling for individuals fixed effects.

The next two columns, (3) and (4), are meant to test asymmetric learning using the strategy suggested by equation (3.5.17). In column (3) the interaction between previous training and tenure is linear and in column (4) quadratic. Training at the current job has a similar estimated effect in the specifications, about 5 percent of the hourly wage, while the coefficient of training at previous jobs is close to zero and not statistically significant. However, the coefficient of the interaction of that variable with tenure is positive and statistically different from zero. After one year at the new job the hourly wage increases by 1 percent if the worker had training within a previous job. Column (4) indicates that the effect of previous training along tenure may be not linear. The returns of training at previous jobs increase with tenure at a decreasing rate.

Columns (5) and (6) are the estimates associated with the second strategy—i.e., using the specification of equation (3.5.18). The only difference between these two columns is that the second one includes the interaction between tenure and TP. Under this strategy it is assumed that there is not adverse selection for workers who are separated from their jobs due to plant closing. The plant closing indicator variable has a negative coefficient in both columns, but in neither of them is statistically significant. The interaction between plant closings and training at the previous job has a positive sign and it is statistically significant but only at a 10 percent confidence level. The magnitude of the coefficient in each specification is larger than the one of training at the current job. However, the standard error of the coefficients of the interaction terms are relatively large. The test with

job sponsored by the worker, government, or other indicators.

Table 3.4: Employer-sponsored training returns
Dependent Variable: Log of Hourly Wages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TC	0.161	0.052	0.051	0.050	0.042	0.043	0.050
	$(0.009)^{***}$	$(0.015)^{***}$	$(0.015)^{***}$	$(0.015)^{***}$	$(0.016)^{***}$	$(0.016)^{***}$	$(0.015)^{***}$
TP	0.116	0.031	0.011	−0.008	0.001	0.015	0.016
	$(0.010)^{***}$	(0.023)	(0.025)	(0.027)	(0.027)	(0.026)	(0.041)
Tenure	0.033	0.008	0.007	0.007	0.006	0.007	0.007
	$(0.003)^{***}$	$(0.004)^{**}$	$(0.004)^{*}$	$(0.004)^{*}$	(0.004)	$(0.004)^{*}$	$(0.004)^{*}$
TP × *Tenure*			0.010	0.027		0.008	0.010
			$(0.004)^{**}$	$(0.009)^{***}$		$(0.005)^{*}$	$(0.004)^{**}$
TP × *Tenure*^{2}				−0.002			
				$(0.001)^{**}$			
Plant Closing					−0.007	−0.007	
					(0.025)	(0.025)	
TP × *Plant Closing*					0.114	0.115	
					$(0.060)^{*}$	$(0.059)^{*}$	
Experience	0.021	0.067	0.067	0.066	0.067	0.067	0.067
	$(0.005)^{***}$	$(0.012)^{***}$	$(0.012)^{***}$	$(0.012)^{***}$	$(0.013)^{***}$	$(0.013)^{***}$	$(0.012)^{***}$
TP × *Experience*							−0.001
							(0.003)
Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
F-test							
TC=*TP*	10.6	1.1	3.5	6.7	1.7	3.2	0.8
p-value	*0.001*	*0.289*	*0.060*	*0.010*	*0.190*	*0.073*	*0.385*
Observations	23,343	23,343	23,343	23,343	21,837	21,837	23,343
Individuals	9,059	9,059	9,059	9,059	8,871	8,871	9,059

Source: Own calculations based on microdata from Social Protection Survey (SPS) 2002-2015.
Note: Robust standard errors in parentheses in (1) and cluster standard errors at the individual level in parentheses in columns (2) to (7). *TC*=1 if training with current employer and 0 otherwise, *TP*=1 if training with previous employer and 0 otherwise, All equations control for an intercept, tenure, tenure squared, experience, experience squared, age at the beginning of the job, part-time job, informal labor, number of previous jobs, union status, first job, marital status, region, type of employment (permanent, temporal, fixed time, by task or other), economic activity, occupation, survey year, accumulated time on unemployment, inactive and on self-employment, if participated in emergency employment program, transitions between economic activities and occupations, accumulated time on each economic activity and occupation, separation reason in previous jobs, attend educational institution and training in current and previous job sponsored by the worker, government or other. Sample includes salary workers aged between 22 and 65 years old. Individuals for whom there is no information before they were 20 years old are not included. ***, **, * statistical significance at the 1%, 5%, and 10% level, respectively.

the null hypothesis that the coefficient of TC is the same as the one of the interactions cannot be rejected ($F = 1.40$, *p-value = 0.237*, not reported in the table). These results provide additional evidence of asymmetric learning.

Finally, column (7) presents a different specification. It includes an interaction term between experience and training with previous employer. If the coefficient were positive, it could suggest that the relationship between tenure and TP is due to different reasons than asymmetric learning. However, the coefficient is small and not statistically different from zero.

In summary, these results suggest that firms learn training quality asymmetrically. Training that takes place with the current employer seems to have a return of 5 percent of the hourly wage, but workers lose these returns when they change jobs. And as firms learn their productivity, workers get those returns back. The second panel of Table 3.4 presents *F-tests* that evaluate if the coefficients of TC and TP are statistically different. The null hypothesis is rejected in specifications reported in columns (1), (3), (4), and (6). That is, every time the interaction between tenure and TP is included (except for column 7), supporting the previous description.

Lastly, Table 3.5 presents specification including interactions between tenure and the indicator variables of training with the current and previous employer. The interaction between tenure and training with current employer is small and not statistically significant. This result is in line with the prediction of the model for the difference of the evolution of the wage return depending on if the worker participated in the training program with the current or previous employer.

3.7 Conclusions

Several papers have argued and empirically tested for the presence of employers asymmetric learning. They focus on uncertainty about workers' innate ability not considering possible changes in productivity over time. The consequence of this phenomenon involves inefficiencies in many outcomes, such as wages, job mobility, job assignment, and human capital accumulation. However, these issues should disappear over time if, at least to some

Table 3.5: On-the-Job training returns and tenure

Dep. variable: Log of Hourly Wages	(1)	(2)	(3)
TC	0.053	0.052	0.051
	(0.018)***	(0.018)***	(0.021)**
TP	0.030	0.010	−0.010
	(0.023)	(0.025)	(0.027)
Tenure	0.008	0.007	0.007
	(0.004)**	(0.004)*	(0.004)*
$TC \times Tenure$	−0.001	−0.001	−0.001
	(0.002)	(0.002)	(0.004)
$TC \times Tenure^2$			−0.000
			(0.000)
$TP \times Tenure$		0.010	0.027
		(0.004)**	(0.009)***
$TP \times Tenure^2$			−0.002
			(0.001)**
Fixed Effects	Yes	Yes	Yes
F-test			
$TC = TP$	1.1	3.1	4.9
p-value	*0.302*	*0.078*	*0.027*
$TC\times Tenure = TP\times Tenure$		5.2	7.7
p-value		*0.022*	*0.005*
$TC\times Tenure^2 = TP\times Tenure^2$			5.5
p-value			*0.019*
Observations	23,342	23,342	23,342
Individuals	9,059	9,059	9,059

Source: Own calculations based on microdata from Social Protection Survey (SPS) 2002-2015. Note: Cluster standard errors at individual level in parenthesis. Sample includes salary workers aged between 22 and 65 years old. Individuals for whom there is no information before they were 20 years old are not included. *TC*=1 if training with current employer and 0 otherwise, *TP*=1 if training with previous employer and 0 otherwise, Same controls included than specification in presented in Table 3.4. ***, **, * statistical significance at the 1%, 5%, and 10% level, respectively.

degree, there is symmetric learning. Nevertheless, if the information asymmetry exists for any changes in productivity further inefficiencies still arise.

This chapter considers asymmetric learning about the productivity returns of on-the-job training programs. This chapter's contribution is twofold: First, I build a model where the information asymmetry comes from changes in workers' productivity with testable implications regarding the way information is transmitted across employers. Moreover, I perform the tests on the Chilean Social Protection Survey.

I use the Chilean Social Protection Survey to empirically test the predictions of the model using a fixed effects estimator for the returns of on-the-job training on hourly wages. The results suggest that the training firm has more information than potential employers. Asymmetric learning regarding the productivity gains of OTJ training can lead to inefficiency costs in job allocation and human capital accumulation. Policies that aim to provide information regarding different training programs and their benefits can help ameliorate these problems.

References

Acemoglu, D., and J.-S. Pischke. 1998. Why Do Firms Train? Theory and Evidence. *The Quarterly Journal of Economics* 113(1):79–119.

Agencia de Calidad de la Educación. 2015. Informe Técnico SIMCE. Tech. Rep.

Agostinelli, Francesco. 2018. Investing in Children's Skills: An Equilibrium Analysis of Parental Investments and Children's Social Interaction. *Unpublished Manuscript*.

Agostinelli, Francesco, Morteza Saharkhiz, and Matthew Wiswall. 2020. Home and School in the Development of Children. *NBER Working Paper Series* No. 26037.

Agostinelli, Francesco, and Matthew Wiswall. 2016. Estimating the Technology of Children's Skill Formation. *NBER Working Paper Series* No. 22442.

Albornoz, Facundo, Samuel Berlinski, and Antonio Cabrales. 2018. Motivation, resources, and the organization of the school system. *Journal of the European Economic Association* 16(1):199–231.

Altonji, Joseph G., and Charles R. Pierret. 2001. Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics* 116(1):313–350.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. 2012. Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics* 3(3):421–470.

Attanasio, Orazio, Teodora Boneva, and Christopher Rauh. 2018. Parental Beliefs about Returns to Different Types of Investments in School Children. *HCEO Working Paper Series* 2018-032.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020a. Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. *American Economic Review* 110(1):48–85.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2020b. Human Capital Development and Parental Investment in India. *The Review of Economic Studies* 026:1–31.

Ballou, Dale. 2009. Test Scaling and Value-Added Measurement. *Education Finance and Policy* 4(4):351–383.

Barron, John M., Dan A. Black, and Mark A. Loewenstein. 1989. Job Matching and On-the-Job Training. *Journal of Labor Economics* 7(1):1–19.

Behrman, Jere, Susan Parker, Petra Todd, and Kenneth I. Wolpin. 2015. Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment In Mexican High Schools. *Journal of Political Economy* 123(2):325–364.

Behrman, Jere R., Michela M. Tincani, Petra E. Todd, and Kenneth I. Wolpin. 2016. Teacher quality in public and private schools under a voucher system: The case of Chile. *Journal of Labor Economics* 34(2):319–362.

Berniell, I., and R Estrada. 2020. Poor Little Children: The Socioeconomic Gap in Parental Responses to School Disadvantage. *Labour Economics* 66(23).

Bharadwaj, Prashant, Juan Eberhard, and Christopher Neilson. 2018. Health at Birth, Parental Investments, and Academic Outcomes. *Journal of Labor Economics* 36(2):349–394.

Bock, R. Darrell, and Murray Aitkin. 1981. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* 46(4):443–459.

Caetano, Gregorio, Josh Kinsler, and Hao Teng. 2019. Towards causal estimates of children's time allocation on skill development. *Journal of Applied Econometrics* 34(4):588–605.

Carneiro, Pedro, and James J. Heckman. 2003. Human Capital Policy. In *Inequality in america: What role for human capital policies*, ed. James J. Heckman and Alan Krueger. Cambridge, MA: MIT Press.

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

Chang, Kang-Hung, Deborah A. Cobb-Clark, and Nicolas Salamanca. 2020. Parents' Responses to Teacher Qualifications. *IZA Discussion Papers* No. 13065.

Chetty, By Raj, John N Friedman, and Jonah E Rockoff. 2014a. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* 104(9):2593–2632.

———. 2014b. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review* 104(9):2633–2679.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from project star. *Quarterly Journal of Economics* 126(4): 1593–1660.

Cooper, Harris, Jorgianne Civey Robinson, and Erika A Patall. 2006. Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003. *Review of Educational Research* 76(1):1–62.

Cunha, Flavio, Irma Elo, and Jennifer Culhane. 2013. Eliciting maternal expectations about the technology of cognitive skill formation. *NBER Working Paper Series* No. 19144.

Cunha, Flavio, and James J. Heckman. 2008. Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources* 43(4):738–782.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica* 78(3):883–931.

Cunha, Flavio, Lance Lochner, J Heckman, and Dimitry V Masterov. 2006. Interpreting the Evidence on Life Cycle Skill Formation. In *handbook of the economics of education*, ed. Eric A. Hanushek and Frank Welch, 697–812. 11331, Amsterdam: North-Holland: Elsevier.

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics* 5(2):29–57.

Datar, Ashlesha, and Bryce Mason. 2008. Do reductions in class size "crowd out" parental investment in education? *Economics of Education Review* 27(6):712–723.

De Fraja, Gianni, Tania Oliveira, and Luisa Zanchi. 2010. Must try harder: Evaluating the role of effort in educational attainment. *Review of Economics and Statistics* 92(3):577–597.

Del Boca, Daniela, Christopher Flinn, Ewout Verriest, and Matthew Wiswall. 2019. Actors in the Child Development Process. *HCEO Working Paper Series* 2019-005.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. Household Choices and Child Development. *Review of Economic Studies* 81(1):137–185.

———. 2016. Transfers to Households with Children and Child Development. *The Economic Journal* 126(1):136–183.

Deming, David. 2009. Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics* 1(3):111–34.

Farber, Henry S, and Robert Gibbons. 1996. Learning and wage dynamics. *Quarterly Journal of Economics* 111(4):1007–1047.

Fiorini, Mario, and Michael P. Keane. 2014. How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics* 32(4):787–836.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2016. Parental responses to public investments in children: Evidence from a maximum class size rule. *Journal of Human Resources* 51(4):832–868.

Fu, Chao, and Nirav Mehta. 2018. Ability Tracking, School and Parental Effort, and Student Achievement: A Structural Model and Estimation. *Journal of Labor Economics* 36(4):923–979.

Gensowski, Miriam, Rasmus Landersø, Dorthe Bleses, Philip Dale, Anders Højen, and Laura Justice. 2020. Public and Parental Investments and Children's Skill Formation. *Unpublished manuscript*.

Gibbons, Robert, and Lawrence F. Katz. 1991. Layoffs and Lemons. *Journal of Labor Economics* 9(4):351–380.

Golan, Limor. 2005. Counteroffers and Efficiency in Labor Markets with Asymmetric Information. *Journal of Labor Economics* 23(2):373–393.

Goldhaber, Dan, and Roddy Theobald. 2013. Managing the teacher workforce in austere times: The determinants and implications of teacher layoffs. *Education Finance and Policy* 8(4):494–527.

Greaves, Ellen, Iftikhar Hussain, Birgitta Rabe, and Imran Rasul. 2019. Parental Responses to Information About School Quality: Evidence from Linked Survey and Administrative Data. *ISER Working Paper Series* No. 2019-0.

Greenwald, Bruce C. 1986. Adverse selection in the labour market. *Review of Economic Studies* 53(3):325–347.

Hanushek, Eric A. 2011. The economic value of higher teacher quality. *Economics of Education Review* 30(3):466–479.

———. 2020. Education production functions. In *Handbook of education economics: A comprehensive overview*, ed. Steve Bradley and Colin Green, 2nd ed., chap. 13, 161–170. London: Academic Press.

Hanushek, Eric A, and Steven G Rivkin. 2012. The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics* 4(1):131–157.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz. 2010. Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics* 1(1):1–46.

Houtenville, Andrew J., and Karen S. Conway. 2008. Parental Effort, School Resources, and Student Achievement. *Journal of Human Resources* 43(2):437–453.

Hu, Luojia, and Christopher Taber. 2011. Displacement, Asymmetric Information, and Heterogeneous Human Capital. *Journal of Labor Economics* 29(1):113–152.

Jackson, Kirabo C., Jonah E. Rockoff, and Douglas Staiger. 2014. Teacher Effects and Teacher-Related Policies. *Annual Review of Economics* 6(1):801–825.

Jacob, Brian, and Jesse Rothstein. 2016. The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives* 30(3):85–107.

Jacqz, Irene. 2020. The cognitive cost of toxic chemicals in early childhood. *Unpublished Manuscript*.

Johnson, Rucker C., and C. Kirabo Jackson. 2019. Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal: Economic Policy* 11(4):310–349.

Junker, Brian, Lynne Steuerle Schofield, and Lowell J. Taylor. 2012. The use of cognitive ability measures as explanatory variables in regression analysis. *IZA Journal of Labor Economics* 1(1):1–19.

Kahn, Lisa B. 2013. Asymmetric information between employers. *American Economic Journal: Applied Economics* 5(4):165–205.

Koedel, Cory, Kata Mihaly, and Jonah E Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* 47:180–195.

Lentz, Rasmus, and Nicolas Roys. 2015. Training and Search On the Job.

Levitt, Steven, John List, and Sally Sadoff. 2016. The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment. *NBER Working Paper* No. 22107.

Loewenstein, Mark A., and James R. Spletzer. 1998. Dividing the Costs and Returns to General Training. *Journal of Labor Economics* 16(1):142–171.

Lord, Frederic M. 1980. *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbanm.

———. 1983. Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika* 48(2):233–245.

Mansfield, Richard K. 2015. Teacher Quality and Student Inequality. *Journal of Labor Economics* 33(3):751–788.

Ministerio Trabajo y Previsión Social. 2017. *Encuesta de Protección Social*. Santiago de Chile, Chile.

Mislevy, Robert J. 1991. Randomization-Based Inference About Latent Variables From Complex Samples: The Case of Two-Stage Sampling. *Psychometrika* 56(2):177–196.

Mislevy, Robert J. 1993. Should "multiple imputations" be treated as "multiple indicators"? *Psychometrika* 58(1):79–85.

Neal, Derek. 2011. The Design of Performance Pay in Education. In *Handbook of the economics of education*, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, vol. 4, 495–550. Amsterdam: North-Holland: Elsevier.

Neilson, Christopher. 2013. Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students. *Unpublished manuscript*.

Nicoletti, Cheti, and Valentina Tonei. 2020. Do parental time investments react to changes in child's skills and health? *European Economic Review* 127:103491.

Parent, Daniel. 1999. Wages and Mobility: The Impact of Employer-Provided Training. *Journal of Labor Economics* 17(2):298–317.

———. 2003. Employer-supported training in Canada and its impact on mobility and wages. *Empirical Economics* 28(3):431–459.

Pinkston, Joshua C. 2009. A model of asymmetric employer learning with testable implications. *Review of Economic Studies* 76(1):367–394.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. Going to a Better School: Effects and Behavioral Responses. *American Economic Review* 103(4):1289–1324.

Rabe, Birgitta. 2020. Schooling inputs and behavioral responses by families. In *Handbook of education economics: A comprehensive overview*, ed. Steve Bradley and Colin Green, 2nd ed., chap. 16, 217–227. London: Academic Press.

Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rueda, M. Rosario, Jin Fan, Bruce D. McCandliss, Jessica D. Halparin, Dana B. Gruber, Lisha Pappert Lercari, and Michael I. Posner. 2004. Development of attentional networks in childhood. *Neuropsychologia* 42(8):1029–1040.

Samejima, Fumi. 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement* (17).

Samejima, Fumiko. 1993. The bias function of the maximum likelihood estimate of ability for the dichotomous response level. *Psychometrika* 58(2):195–209.

San Martín, Ernesto, Jean Marie Rolin, and Luis M. Castro. 2013. Identification of the 1PL Model with Guessing Parameter: Parametric and Semi-parametric Results. *Psychometrika* 78(2):341–379.

Schofield, Lynne Steuerle. 2014. Measurement error in the AFQT in the NLSY79. *Economics Letters* 123(3):262–265.

Schofield, Lynne Steuerle, Brian Junker, Lowell J. Taylor, and Dan A. Black. 2015. Predictive Inference Using Latent Variables with Covariates. *Psychometrika* 80(3):727–747.

Schönberg, Uta. 2007. Testing for Asymmetric Employer Learning. *Journal of Labor Economics* 25(4):651–691.

Springer, MG, D Ballou, and L Hamilton. 2011. Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT). *Society for Research on Educational Effectiveness*.

Staiger, Douglas O., and Jonah E. Rockoff. 2010. Searching for effective teachers with imperfect information. *Journal of Economic Perspectives* 24(3):97–118.

Steinberg, Laurence, Sandra Graham, Lia O Brien, Jennifer Woolard, and Marie Banich. 2009. Society for Research in Child Development Age Differences in Future Orientation and Delay Discounting. *Child Development* 80(1):28–44.

Strøm, Bjarne, and Torberg Falch. 2020. The role of teacher quality in education production. In *Handbook of education economics: A comprehensive overview*, ed. Steve Bradley and Colin Green, 2nd ed., chap. 22, 307–319. London: Academic Press.

Tincani, Michela Maria. 2020. Teacher labor markets, school vouchers and student cognitive achievement: Evidence from Chile. *Quantitative Economics*, (*forthcoming*).

Todd, Petra E., and Kenneth I. Wolpin. 2003. On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal* 113(1):3–33.

———. 2007. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital* 1(1):91–136.

Waldman, Michael. 1984. Job Assignments, Signalling, and Efficiency. *The RAND Journal of Economics* 15(2):255–267.

———. 1990. Up-or-Out Contracts: A Signaling Perspective. *Journal of Labor Economics* 8(2):230–250.

Wang, Jinwen. 2020. What Drives Inequality in Parental Investment in Children? *Unpublished manuscript*.

Warm, Thomas A. 1989. Weighted likelihood estimation of ability in item response theory. *Psychometrika* 54(3):427–450.

Williams, Benjamin. 2019. Controlling for ability using test scores. *Journal of Applied Econometrics* 34(4):547–565.

Zhang, Ye. 2007. Employer Learning Under Asymmetric Information : The Role of Job Mobility. *Unpublished paper*.

Appendix

Appendix A Data

3.7.1 Chilean Time Use Survey 2015

The Time Use Survey of Chile was collected in 2015. It provides information on the activity diaries for the last week and weekend day of each individual in the sample. I identify the activities parents do with their children, such as helping with studying, doing homework, or any other activity, and the time children spend doing homework, studying or doing other academic activity outside of school. From this information, I estimate the empirical distribution of parental time and children time self-investment.

3.7.2 Survey of Wechsler Intelligence Scale for Children (WISC-V)

The survey of the Wechsler Intelligence Scale for Children (WISC-V) test was collected by the Center for the Development of Inclusion Technologies (CEDETi UC) of the School of Psychology of the Pontifical Catholic University of Chile (PUC). This survey was collected in 2017 and it is nationally representative of the population of children between 6 and 16 years old. The WISC-V test takes 45-65 minutes to administer. The test is designed to generate a full scale IQ measure that represents a child's general intellectual ability. It consists of a set of fifteen different cognitive test. These subtests are labeled as: block design, similarities, matrix reasoning, digit span, coding, vocabulary, picture completion, picture concepts, symbol search, information, letter-number sequencing, cancellation, comprehension, arithmetic, word reasoning.

Appendix B: Time investments measurement

3.7.3 Results identification measurement system

The identification of the parameters of the response model follows San Martín et al. (2013).
Proof that $q(\cdot)$ is increasing in β_s. Let f and F be the density and cumulative distribution of the logistic distribution, respectively, and assume $S \geq 3$ and $K_s = 2$ for all s.

$$p_{s,s'} = \int F(\beta_s h - \alpha(\beta_s, p_s))F(\beta_{s'} h - \alpha(\beta_{s'}, p_{s'}))g(h)dh \equiv q(\beta_s, p_s, \beta_{s'}, p_{s'}) \tag{3.7.1}$$

Taking the derivative of equation (3.7.1) with respect to β_s and rearranging:

$$\frac{\partial q(\beta_s, p_s, \beta_{s'}, p_{s'})}{\partial \beta_s} = \int f(\beta_s h - \alpha(\beta_s, p_s))(h - \frac{\partial \alpha}{\partial \beta_s}(\beta_s, p_s))F(\beta_{s'} h - \alpha(\beta_{s'}, p_{s'}))g(h)dh \tag{3.7.2}$$

Replace $\alpha(\beta_s, p_s)$ in equation (1.5.7) and take the derivative with respect to β_1 and rearrange:

$$\frac{\partial \alpha}{\partial \beta_s}(\beta_s, p_s) = \int h g_C(h; \beta_s, p_s)dh = \mathrm{E}_{\beta_s, \alpha(\beta_s, p_s)}[h] \tag{3.7.3}$$

where

$$g_C(h; \beta_s, p_s) = \frac{f(\beta_s h - \alpha(\beta_s, p_s))g(h_i)}{\int f(\beta_s h - \alpha(\beta_s, p_s))g(h)dh} \tag{3.7.4}$$

Multiply and divide equation (3.7.2) by $\int f(\beta_s h - \alpha(\beta_s, p_s)g(h)dh$, we can rewrite the equation as:

$$\frac{\partial q(\beta_s, p_s, \beta_{s'}, p_{s'})}{\partial \beta_s} = \int (h - \mathrm{E}_{\beta_s, \alpha(\beta_s, p_s)}[h])F(\beta_{s'} h - \alpha(\beta_{s'}, p_{s'}))g_C(h; \beta_s, p_s)dh \tag{3.7.5}$$

Then, define:

$$\mathrm{E}_{\beta_s,\alpha(\beta_s,p_s)}[F(\beta_{s'}h-\alpha(\beta_{s'},p_{s'}))] = \int F(\beta_{s'}h-\alpha(\beta_{s'},p_{s'})g_C(h;\beta_s,p_s)dh \tag{3.7.6}$$

Subtracting $\mathrm{E}_{\beta_s,p_s}[F(\beta_{s'}h-\alpha(\beta_{s'},p_{s'}))]\int(h-\mathrm{E}_{\beta_s,p_s}[h])g_C(h;\beta_s,p_s)dh = 0$ on left side of equation (3.7.5) and rearranging:

$$\begin{aligned}
&\frac{\partial q(\beta_s,p_s,\beta_{s'},p_{s'})}{\partial \beta_s} = \\
&\quad \int (h-\mathrm{E}_{\beta_s,\alpha(\beta_s,p_s)}[h])(F(\beta_{s'}h-\alpha(\beta_{s'},p_{s'}))-\mathrm{E}_{\beta_s,\alpha(\beta_s,p_s)}[F(\beta_{s'}h-\alpha(\beta_{s'},p_{s'}))])g_C(h;\beta_s,p_s)dh \\
&\qquad\qquad \times \int f(\beta_s h-\alpha(\beta_s,p_s)g(h_i)dh_i \\
&\quad = \mathrm{Cov}_{\beta_s,\alpha(\beta_s,p_s)}(h, F[\beta_{s'}h-\alpha(\beta_{s'},p_{s'})]) \times \int f(\beta_s h-\alpha(\beta_s,p_s)g(h)dh > 0
\end{aligned} \tag{3.7.7}$$

This is positive since $\beta_{s'} > 0$ and so the covariance between h and $F(\beta_{s'}h-\alpha(\beta_{s'},p_{s'}))$ is positive as well.

Proof that $r(\beta_1,p_1,p_2,p_3,p_{1,2},p_{1,3})$ is strictly decreasing in β_1.

$$\begin{aligned}
&\frac{\partial r}{\partial \beta_1}(\beta_1,p_1,p_2,p_3,p_{1,2},p_{1,3}) = \\
&\qquad \frac{\partial q}{\partial \beta_2}(\overline{q}(\beta_1,p_1,p_{1,2},p_2),p_2,\overline{q}(\beta_1,p_1,p_{1,3},p_3),p_3) \times \frac{\partial \overline{q}}{\partial \beta_1}(\beta_1,p_1,p_{1,2},p_2) \\
&\qquad + \frac{\partial q}{\partial \beta_3}(\overline{q}(\beta_1,p_1,p_{1,2},p_2),p_2,\overline{q}(\beta_1,p_1,p_{1,3},p_3),p_3) \times \frac{\partial \overline{q}}{\partial \beta_1}(\beta_1,p_1,p_{1,3},p_3)
\end{aligned} \tag{3.7.8}$$

Since $\partial q/\partial \beta_s > 0$ for all s and given that:

$$q(\beta_1,p_1,\overline{q}(\beta_1,p_1,p_{1,2},p_2),p_2) = p_{1,2} \tag{3.7.9}$$

take the derivative with respect to β_1

$$\frac{\partial q}{\partial \beta_1}(\beta_1, p_1, \overline{q}(\beta_1, p_1, p_{1,2}, p_2), p_2) + \frac{\partial q}{\partial \beta_2}(\beta_1, p_1, \overline{q}(\beta_1, p_1, p_{1,2}, p_2), p_2) \times \frac{\partial \overline{q}}{\partial \beta_1}(\beta_1, p_1, p_{1,2}, p_2) = 0 \tag{3.7.10}$$

Rearranging terms and since $\partial q / \partial \beta_s > 0$ for all s we have:

$$\frac{\partial \overline{q}}{\partial \beta_1}(\beta_1, p_1, p_{1,2}, p_2) = -\frac{\dfrac{\partial q}{\partial \beta_1}(\beta_1, p_1, \overline{q}(\beta_1, p_1, p_{1,2}, p_2), p_2)}{\dfrac{\partial q}{\partial \beta_2}(\beta_1, p_1, \overline{q}(\beta_1, p_1, p_{1,2}, p_2), p_2)} < 0 \tag{3.7.11}$$

Then, since $\partial q / \partial \beta_s > 0$ and $\partial \overline{q} / \partial \beta_s < 0$ for all s, it follows from equation (3.7.8) that $\partial r / \partial \beta_1 < 0$.

3.7.4 Functional forms

As stated in the main text, the time investment variables observed in the administrative data are denoted by Z_{its}, where s indexes the measure. These are ordered categorical questions—i.e., $Z_{its} \in \{1, 2, \ldots, K_s\}$ where K_s denotes the number of categories of measure s. I assume a multivariate ordered response model with a latent factor, h_{it} (time investment). Let S_t be the number of ordered categorical questions at grade t. The Fisher information function is:

$$\mathrm{I}(h_{it}) = \sum_{s=1}^{S_t} \sum_{k=1}^{K_s} \frac{\left(\dfrac{\partial \mathrm{Pr}_{sk}}{\partial h_{it}}\right)^2 - \mathrm{Pr}_{sk} \dfrac{\partial^2 \mathrm{Pr}_{sk}}{\partial h_{it}^2}}{\mathrm{Pr}_{sk}^2}. \tag{3.7.12}$$

where $\mathrm{Pr}_{sk} = \mathrm{Pr}(Z_{its} = k \mid h_{it})$. Lord (1983) provides a specification for the bias function of the maximum likelihood estimator in the dichotomous measure setting, while Samejima (1993) extends the result for the polytomous case. See next section for the definition of this estimator. This corresponds to the function $B(\cdot)$ in equation (1.5.13) of Section 1.5.1. This function is:

$$\text{Bias}(h_{it}) = -\frac{1}{2[I(h_{it})]^2} \sum_{s=1}^{S_t} \sum_{k=1}^{K_s} \frac{\dfrac{\partial \text{Pr}_{sk}}{\partial h_{it}} \dfrac{\partial^2 \text{Pr}_{sk}}{\partial h_{it}^2}}{\text{Pr}_{sk}}. \tag{3.7.13}$$

3.7.5 Alternative estimators of time investments

Expected a posteriori Estimator

With the estimates of (α, β), it is possible to compute the posterior distribution of time investment—i.e., the distribution of h_{it} conditional on the responses to ordered categorical questions of a particular student, $\mathbf{z}_{it}$. This distribution tells us how likely the student is to have each level of time investment conditional on the responses to these ordered categorical question.

$$\tilde{g}_t(h_{it} \mid \mathbf{z}_{it}; \alpha, \beta) = \frac{f_t(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta) g_t(h_{it})}{\int f(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta) g_t(h_{it}) dh_{it}} \tag{3.7.14}$$

The posterior expected value estimator is defined as follows:

$$h_{it}^{EAP} = \text{E}(h_{it} \mid \mathbf{z}_{it}) = \int h_{it} \tilde{g}(h_{it} \mid \mathbf{z}_{it}; \alpha, \beta) dh_{it} \tag{3.7.15}$$

Other *Bayesian* estimators are the mode or median of $\tilde{g}(h_{it} \mid \mathbf{z}_{it}; \alpha, \beta)$ and the same properties apply to them.

Maximum Likelihood Estimator

Another estimator for each student's time investment is the value that maximizes the individual likelihood. That is,

$$h_{it}^{ML} \equiv \underset{h_{it}}{\text{argmax}} \log f(\mathbf{z}_{it} \mid h_{it}; \alpha, \beta) \tag{3.7.16}$$

The bias of this estimator is $\text{Bias}(h_{it})$ in equation (3.7.13).

Appendix C: Monte Carlo of time measurement systems

This appendix presents a Monte Carlo exercise to compare the performance of linear and non-linear latent factor models using ordered categorical questions. The latent factor system is used to estimate parameter of the structural relationship between the latent factor (time investment) and additional variables—e.g., classroom inputs and skills. Data generating process: Let the classroom inputs be $C \sim N(0,1)$, and the time investment policy function:

$$h = \delta C + \eta \tag{3.7.1}$$

where $\eta \sim N(0, \sigma_\eta^2)$ and without loss of generality $\sigma_\eta = \sqrt{(1-\delta^2)}$. So that $h \sim N(0,1)$. Skills, labeled by θ, have a production technology with a single input h:

$$\theta = \gamma h + \nu \tag{3.7.2}$$

where $\nu \sim N(0,1)$ is an error term. Time investment h is the latent variable and its distribution its known. The parameters of interest are δ and γ.

Time investment is imperfectly measured by the variables Z_s, where s indexes the measure. These are ordered categorical questions—i.e., $Z_s \in \{1, 2, \ldots, K_s\}$ where K_s denotes the number of categories of measure s. In this exercise $K_s = 4$ for all s. These measures are generated with a multivariate ordered response model:

$$Z_s = k \quad \text{if and only if} \quad \alpha_{sk} \leq \beta_s h + \epsilon_s < \alpha_{sk+1} \qquad \text{for } k = 1, 2, \ldots, K_s, \tag{3.7.3}$$

where $\alpha_{s1} = -\infty$ and $\alpha_{sK_s+1} = \infty$. The parameters of each ordered categorical questions are drawn from the following distributions: $\beta_s \sim$ Uniform$(0, 2)$ and $\alpha_{s\tilde{k}} \sim N(0, \beta_s^2)$ with $\tilde{k} = 1, \ldots, K_s - 1$. The $K_s - 1$ realizations of $\alpha_{s\tilde{k}}$ are then ordered in increasing values and the index $\tilde{k}$ is replace by the corresponding ordered index $k = 1, \ldots, K_s - 1$.

I perform the simulation under several assumptions of the distributions of the errors ϵ_s. However, the non-linear model estimation assumes logistic error terms. I perform the simulations under different distribution to evaluate potential bias of miss-specification under non-linear systems.

In the literature, researchers have used linear and non-linear latent factor models for household investments. Linear models treat the ordered categorical variables as continuous variables and assume the following structure:

$$Z_s = \tilde{\alpha}_s + \tilde{\beta}_s h + \varepsilon_s, \tag{3.7.4}$$

under the assumption that $\varepsilon_s \perp \varepsilon_{s'}$ for all $s \neq s'$ and $h \perp \varepsilon_s$ for all s. The system is identified exploiting these orthogonality conditions.

Instead, the non-linear models follow a similar identification strategy than the one presented in Section 1.5.1. Estimation procedures used are the weighted maximum likelihood (WML) estimator described section Section 1.6.1, and maximum likelihood (ML) and expected *a posteriori* (EP) estimators lay out in Appendix B.

I simulate a sample of size N, M times. The Monte Carlo exercise evaluates the asymptotic properties of estimators of δ and γ using different measurement models under different data generating processes. For each simulation, I estimate both the linear and non-linear measurement models. I generate estimates of h for each individual under different estimators. For the linear model:

$$h_s^L = \frac{Z_s - \tilde{\alpha}_s}{\tilde{\beta}_s} \tag{3.7.5}$$

For the non-linear model, I estimate each individual h using the WML, ML, and EP estimators, and labeled them as h_w^E, where $E = \text{WML}, \text{ML}, \text{EP}$ and w is an index of the set of measures used in the estimation. Disjoint sets of the categorical ordered questions generate different measures that provide exclusion conditions for the 2SLS estimator. Table C.1 presents the results.

Table C.1:
Monte Carlo Simulation

measurement error ϵ_s distribution	$\delta = 0.5$				$\gamma = 1$			
	Linear	Non-linear			Linear	Non-linear		
	OLS	EP-OLS	ML-OLS	WML-OLS	IV	EP-OLS	ML-OLS	WML-IV
logistic(0,1)	0.496	0.482	0.511	0.488	0.981	1.008	0.945	0.995
	(0.095)	(0.017)	(0.019)	(0.018)	(0.089)	(0.029)	(0.030)	(0.034)
uniform(-3,3)	0.488	0.479	0.510	0.485	0.976	1.009	0.943	0.998
	(0.105)	(0.016)	(0.019)	(0.018)	(0.174)	(0.029)	(0.032)	(0.036)
normal(0,1)	0.495	0.481	0.505	0.481	0.965	1.013	0.956	1.016
	(0.045)	(0.016)	(0.019)	(0.020)	(0.094)	(0.038)	(0.035)	(0.053)
degenerate 0	0.493	0.481	0.510	0.484	0.948	1.022	0.936	1.067
	(0.036)	(0.099)	(0.096)	(0.103)	(0.100)	(0.230)	(0.157)	(0.338)

Note: Monte Carlo simulation. Standard deviations in parentheses. All non-linear model are estimated under the assumption than $\epsilon_s \sim$ logistic(0,1). Estimators of each individual h: weighted maximum likelihood (WML), maximum likelihood (ML), expected *a posteriori* (EP). Estimators of γ and δ: ordinary least squares (OLS) and two-stage least squares (2SLS).

I estimate γ with 2SLS estimator and δ with OLS estimator. Treating a categorical variable as continuous "adds" measurement error. However, the 2SLS estimator of γ deals with this measurement error and it converges to its true value. Similarly, the 2SLS estimator of γ using the non-linear model is consistent as well. However, under the logistic distribution—i.e., assuming the correct model—the linear model has standard deviation 5 times larger than the non-linear case.

The main difference comes from the estimation of δ. When the noisy measure is on the left side of the equation is not possible to use instrumental variables. Either way, if the measurement error is i.i.d., it implies a precision cost but not bias. However, as mentioned, the continuity assumption over a categorical variable could result in additional measurement error.

Appendix D: Model Solution

I estimate the model using using an indirect inference estimator. This is a simulation based method and it requires to simulate, given the initial conditions, the choices of household at each school grade. Let $\Omega_{it} = \{\theta_{it}, C_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\}$ be the state space vector, where the state variables are child's skill θ_{it}, classroom inputs C_{it}, a vector of demographic characteristics $\mathbf{x}_{it}$ (including household income y_{it}) and other observable inputs $\mathbf{z}_{it}$ in the skill formation. The classroom and income processes—known by the household—are $\log C_{it} = \kappa_t' \mathbf{x}_{it} + \Delta_{it}$ and $\log y_{it} = \overline{y}_t + \rho_t \log y_{it} + \omega_{it}$. When household make decisions they know Δ_{it} and ω_{it}, but they don't know their in subsequent grades or the current skill shock ν_{it}—i.e., the skill technology shock. The household makes expectations regarding the classroom, skill, and income shocks. There are two continuous control variables (parental and student time) and three variables for which I have to integrate to generate expectations of continuation values, which makes this problem computationally intensive. To decrease the computation burden, I use an interpolation approach of the value function. I use Monte Carlo integration for Δ_{it}, and Gauss–Hermite quadratures for ω_{it} and ν_{it}. I estimate an interpolation function as follows: I take as given the current guess of the preference parameters. First, in period $T = 4$, that is, when the household has a child attending tenth grade. I draw $D = 200$ realizations of Δ_{it} from the assumed distribution. Second, I draw randomly $S = 200$ points of support the rest of the variables in Ω_{it}. Third, I simulate the household choices, calculate the value function, and take the average across the D realizations of Δ_{it} and the Gauss–Hermite quadrature integration of ω_{it} and ν_{it}. The end result is the expected value of the value function. Lastly, I regress that expected value function on the values drawn of Ω_{it} (second order polynomial in continuous variables). That is, this regression is run in a size S sample. The coefficients defined the interpolation function and can be used to predict the expected value of the value function at each point in the support of the state space without solving the problem.

Next, using the interpolation function, I proceed to the same exercise for $T = 3$ and $T = 2$. Once I have the interpolation function at each school grade. I simulate the choices of the households in my sample using their initial conditions.

Appendix E: OJT Model Solution

In this appendix I will solve for the close-form solution of equilibrium wage offers.

At any period t non-trained workers will offered their productivity minus hiring cost plus the net present value of expected profits from future periods. Formally, the outside wage offer at period t is:

$$z_t = \mu - h + \beta \mathrm{E}_{\xi}(\pi(0, z(t+1))). \tag{3.7.1}$$

The separation shock only hit at the beginning of period 2. At every period $t \geq 2$, the incumbent firm faces the same problem and the solution for every t is the same. Then, for $t \geq 2$,

$$\begin{aligned} z_t &= \mu - h + \beta \textstyle\sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1}(\mu - z_\tau) \\ &= \mu - h + \frac{\beta}{1-\beta}(\mu - z_t) \\ \frac{z_t}{1-\beta} &= \frac{\mu}{1-\beta} - h \\ z_t &= \mu - (1-\beta)h, \end{aligned} \tag{3.7.2}$$

Thus, incumbent and outside firms' wage offers to untrained workers will be $z_t = \mu - (1-\beta)h$ for every $t \geq 2$. Instead, for $t = 1$ since with probability ξ the workers is exogeneously separated at the beginning of period 2, we have that $z_1 = \mu - [1 - \beta(1-\xi)]h$.

The offer at period 1 for workers who are not trained are:

$$z_1 = \mu - [1 - \beta(1-\xi)]h. \tag{3.7.3}$$

and $z_t = \mu - (1-\beta)h$ for $t \geq 2$.

For the sake of simplicity, I will assume the distribution of training productivity returns is uniform—i.e., $\alpha \sim Uniform(0, \overline{\alpha})$. Workers who were retained in period 2 will receive offers from outside firms. The offers are:

$$\begin{aligned}
v_2^{R,1} &= \mu + \mathrm{E}(\alpha \mid \alpha \geq \alpha^*) - h + \beta \mathrm{E}(\pi(\alpha, w^{R,1}(3)) \mid \alpha \geq \alpha^*) \\
&= \mu + \mathrm{E}(\alpha \mid \alpha \geq \alpha^*) - h + \beta \mathrm{E}\left[\textstyle\sum_{\tau=0}^{\infty} \beta^{\tau} \left(\mu + \alpha - w_{3+\tau}^{R,1}\right) \,\middle|\, \alpha \geq \alpha^*\right] \\
&= \frac{\mu + \mathrm{E}(\alpha \mid \alpha \geq \alpha^*)}{1-\beta} - h - \frac{\beta v_2^{R,1}}{1-\beta} \qquad (3.7.4) \\
&= \mu + \mathrm{E}(\alpha \mid \alpha \geq \alpha^*) - (1-\beta)h = w_2^{R,1} \\
w_2^{R,1} &= \mu + \frac{\overline{\alpha} + \alpha^*}{2} - (1-\beta)h
\end{aligned}$$

The replacement in the previous third line comes from the fact that outside wage offer in equilibrium will be the same at each period $t \geq 2$. Again, the incumbent firm will match the offer, that is, $w_2^{R,1} = v_2^{R,1}$. Furthermore, it will choose α^* such that the firm is indifferent between retaining the worker. That is,

$$\begin{aligned}
\mu + \alpha^* - w_2^{R,1} + \beta\pi(\alpha^*, w^{R,1}(3)) &= 0 \\
\mu + \alpha^* - \mu - \frac{\overline{\alpha} + \alpha^*}{2} + (1-\beta)h + \beta \sum_{\tau=0}^{\infty} \beta^{\tau} \left(\mu + \alpha^* - w_{3+\tau}^{R,1}\right) &= 0 \qquad (3.7.5) \\
\frac{1}{1-\beta}\left(\frac{\alpha^* - \overline{\alpha}}{2} + (1-\beta)h\right) &= 0
\end{aligned}$$

where the step from the second line to the third comes from the fact that from period 2 outside offer to these group of workers will not change at equilibrium. Thus, $\alpha^* = \overline{\alpha} - 2(1-\beta)h$ and the wage offers of period 2 are

$$\begin{aligned}
w_2^{R,1} &= \mu + \frac{\overline{\alpha} + \alpha^*}{2} - (1-\beta)h \\
&= \mu + \overline{\alpha} - 2(1-\beta)h. \qquad (3.7.6)
\end{aligned}$$

The outside offer to lay off/separated workers are given by the following equation:

$$\begin{aligned}
v_2^{L,1} &= \mu - h + \frac{F(\alpha^*)}{F(\alpha^*) + [1 - F(\alpha^*)]\xi} \times \\
&\qquad \Big[\mathrm{E}(\alpha \mid \alpha < \alpha^*) + \beta \frac{[F(\alpha^*) - F(\alpha^{**})]}{F(\alpha^*)} \mathrm{E}(\pi(\alpha, w^{R,2}(3)) \mid \alpha^{**} \leq \alpha < \alpha^*)\Big] \\
&\quad + \frac{[1 - F(\alpha^*)]\xi}{F(\alpha^*) + [1 - F(\alpha^*)]\xi} \times \Big[\mathrm{E}(\alpha \mid \alpha \geq \alpha^*) + \beta \mathrm{E}(\pi(\alpha, w^{R,2}(3)) \mid \alpha \geq \alpha^*)\Big] \\
&= \mu - h + \frac{\alpha^*}{\alpha^* + [\overline{\alpha} - \alpha^*]\xi} \times \Big[\frac{\alpha^*}{2} + \beta \frac{[\alpha^* - \alpha^{**}]}{\alpha^*} \frac{1}{1-\beta}(\mu + \frac{\alpha^{**} + \alpha^*}{2} - w_3^{R,2})\Big] \\
&\quad + \frac{[\overline{\alpha} - \alpha^*]\xi}{\alpha^* + [\overline{\alpha} - \alpha^*]\xi} \times \Big[\frac{\overline{\alpha} + \alpha^*}{2} + \beta \frac{1}{1-\beta}(\mu + \frac{\overline{\alpha} + \alpha^*}{2} - w_3^{R,2})\Big]
\end{aligned} \tag{3.7.7}$$

The retained workers at period 3 (that were previously laid off) will be poach by outside firms. The outside offers are derived in the following equations:

$$\begin{aligned}
v_3^{R,2} &= \mu - h + \frac{F(\alpha^*) - F(\alpha^{**})}{F(\alpha^*) - F(\alpha^{**}) + [1 - F(\alpha^*)]\xi} \times \\
&\qquad \Big[\mathrm{E}(\alpha \mid \alpha^{**} \leq \alpha < \alpha^*) + \beta \mathrm{E}(\pi(\alpha, w^{R,2}(4)) \mid \alpha^{**} \leq \alpha < \alpha^*)\Big] \\
&\quad + \frac{[1 - F(\alpha^*)]\xi}{F(\alpha^*) - F(\alpha^{**}) + [1 - F(\alpha^*)]\xi} \times \Big[\mathrm{E}(\alpha \mid \alpha \geq \alpha^*) + \beta \mathrm{E}(\pi(\alpha, w^{R,2}(4)) \mid \alpha \geq \alpha^*)\Big] \\
&= \mu - (1-\beta)h + \frac{\alpha^* - \alpha^{**}}{\alpha^* - \alpha^{**} + [\overline{\alpha} - \alpha^*]\xi} \times \Big(\frac{\alpha^{**} + \alpha^*}{2}\Big) \\
&\qquad + \frac{[\overline{\alpha} - \alpha^*]\xi}{\alpha^* - \alpha^{**} + [\overline{\alpha} - \alpha^*]\xi} \times \Big(\frac{\overline{\alpha} + \alpha^*}{2}\Big)
\end{aligned} \tag{3.7.8}$$

where $w_3^{R,2}$ is defined by the outside offer of those workers (who will not be lay off again), essentially the same as the one they would had have if their stayed with the previous employer. New incumbent firm offers $w_3^{R,2} = v_3^{R,2}$. Moreover, it chooses α^{**} such that it is indifferent between retaining and laying off the worker. Formally,

$$\begin{aligned} \mu+\alpha^{**}-w_3^{R,2}+\beta\pi(\alpha^{**},w^{R,2}(4)) &= 0 \\ \mu+\alpha^{**}-w_3^{R,2}+\beta\sum_{\tau=0}^{\infty}\beta^{\tau}\left(\mu+\alpha^{**}-w_{4+\tau}^{R,2}\right) &= 0 \\ \mu+\alpha^{**} &= w_4^{R,2}. \end{aligned} \tag{3.7.9}$$

The step from the second line to the forth one comes from the fact that there are no more lay off or exogeneous separation, the retained workers will receive the same offer at period 3 than any subsequent offer, that is $w_3^{R,2}=w_{4+\tau}^{R,2}$ with $\tau=0,1,\ldots$. Thus, the cut off solve the following equation:

$$\alpha^{**}=-(1-\beta)h+\frac{\alpha^*-\alpha^{**}}{\alpha^*-\alpha^{**}+[\overline{\alpha}-\alpha^*]\xi}\times\left(\frac{\alpha^{**}+\alpha^*}{2}\right)+\frac{[\overline{\alpha}-\alpha^*]\xi}{\alpha^*-\alpha^{**}+[\overline{\alpha}-\alpha^*]\xi}\times\left(\frac{\overline{\alpha}+\alpha^*}{2}\right). \tag{3.7.10}$$

Thus, $\alpha^{**}=\overline{\alpha}-4(1-\xi)(1-\beta)h$ and,

$$\begin{aligned} w_3^{R,2} &= \mu+\overline{\alpha}-4(1-\xi)(1-\beta)h \\ &= v_3^{R,2}. \end{aligned} \tag{3.7.11}$$

The outside offer to lay off workers

$$\begin{aligned} v_3^{L,2} &= \mu-h+\mathrm{E}(\alpha\mid\alpha<\alpha^{**})+\beta\mathrm{E}(\pi(\alpha,w^{R,3}(4))\mid\alpha<\alpha^{**}) \\ &= \mu-h+\frac{\alpha^{**}}{2}+\beta\sum_{\tau=0}^{\infty}\beta^{\tau}\left(\mu+\frac{\alpha^{**}}{2}-w_{4+\tau}^{R,3}\right)=w_4^{R,3} \\ w_4^{R,3} &= \mu+\frac{\alpha^{**}}{2}-(1-\beta)h. \end{aligned} \tag{3.7.12}$$

We can rewrite the offer to laid off workers at period 2,

$$
\begin{aligned}
v_2^{L,1} &= \mu - h + \frac{\alpha^*}{\alpha^* + [\overline{\alpha} - \alpha^*]\xi} \times \left[\frac{\alpha^*}{2} + \beta\frac{[\alpha^* - \alpha^{**}]}{\alpha^*}\frac{1}{1-\beta}\left(\mu + \frac{\alpha^{**} + \alpha^*}{2} - w_3^{R,2}\right)\right] \\
&\quad + \frac{[\overline{\alpha} - \alpha^*]\xi}{\alpha^* + [\overline{\alpha} - \alpha^*]\xi} \times \left[\frac{\overline{\alpha} + \alpha^*}{2} + \beta\frac{1}{1-\beta}\left(\mu + \frac{\overline{\alpha} + \alpha^*}{2} - w_3^{R,2}\right)\right] \\
&= \mu + \frac{\overline{\alpha}}{2} + \beta h + \frac{\overline{\alpha} - 2(1-\beta)h}{\overline{\alpha} - 2(1-\xi)(1-\beta)h} \times \left[\frac{(2(1-\xi)-1)^2 2(1-\beta)h}{\overline{\alpha} - 2(1-\beta)h}\beta h\right] \\
&\quad + \frac{[2(1-\beta)h]\xi}{\overline{\alpha} - 2(1-\xi)(1-\beta)h} \times \left[\frac{\overline{\alpha}}{2} + \Big(4(1-\xi)-1\Big)\beta h\right]
\end{aligned} \tag{3.7.13}
$$

The wage offer at period 1 for workers who will be trained is as follows:

$$
\begin{aligned}
v_1 &= \mu + \mathrm{E}(\alpha) - h + \beta[1 - F(\alpha^*)](1-\xi)\mathrm{E}(\pi(\alpha, w^{R,1}(2)) \mid \alpha \geq \alpha^*) - c \\
&= \mu + \mathrm{E}(\alpha) - h + [1 - F(\alpha^*)](1-\xi)\beta h - c \\
v_1 &= \mu + \frac{\overline{\alpha}}{2} - h + \left[\frac{2(1-\beta)h}{\overline{\alpha}}\right](1-\xi)\beta h - c \\
E_c(v_1) &= \mu + \frac{\overline{\alpha}}{2} - h + \left[\frac{(1-\beta)h}{\overline{\alpha}}\right](1-\xi)\beta h
\end{aligned} \tag{3.7.14}
$$

Supplemental Tables and Figures

Supplemental Tables

Table S.1.1:
SIMCE data

Grade	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
2				**x**	**x**	**x**	**x**			
3										
4	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	
5										
6					**x**	**x**	**x**	**x**		**x**
7										
8			**x**		**x**	**x**	**x**		**x**	
9										
10					**x**		**x**	**x**	**x**	**x**

Note: The **x** states which year-school grade combination the SIMCE database has available information. Note that for every cell there is information in the administrative data of the Chilean educational system. Back to Section 1.3.1.

Table S.1.2:
Examples of time investment questions in the SIMCE data

Measure	Parent or child	Question and answers
Parental time		
1	C	My parents help me with my homework. 1: never, 2: sometimes, 3: often, 4: always.
2	C	My parents help me study. 1: never, 2: sometimes, 3: often, 4: always.
3	C	My parents explain me what I do not understand. 1: never, 2: sometimes, 3: often, 4: always.
4	C	My parents are willing to help me when I have problems with a subject or homework. 1: never, 2: sometimes, 3: often, 4: always.
5	P	You talk with the student about how she/he feels in school. 1: never, 2: a few times in a month, 3: a few times in a week, 4: every or almost every day.
6	P	You help the student with school activities. 1: never, 2: a few times in a month, 3: a few times in a week, 4: every or almost every day.
Child time		
1	C	How many days a week from Monday to Friday do you study or do homework? 1: never, 2: 1 or 2 days a week, 3: 3 or 4 days a week, 4: every day.
2	C	I always do my homework. 1: very false, 2: false, 3: true, 4: very true.
3	C	I read what they ask me in school. 1: never or almost never, 2: 1 or 2 a month, 3: 1 or 2 a week, 4: every day or almost every day.
4	C	I strive to do well in all subjects. 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree.
5	C	I am a person who strives to learn. 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree.
6	C	I strive to get good grades. 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree.

Note: These are examples of time investment questions reported in the parent and student questionnaires in the SIMCE administrative data. Back to Section 1.3.3.

Table S.1.3:
Measurement system of skills: WISC-V survey data

Wechsler Intelligence Scale for Children V Sub-test	Location μ_m^A	Scale λ_m^A	Signal share Age 8	10	12	14	16
Block Design	21.3	3.4	0.25	0.23	0.17	0.27	0.14
Similarities	18.6	4.9	0.59	0.49	0.64	0.80	0.49
Matrix Reasoning	15.1	2.8	0.44	0.62	0.68	0.77	0.41
Digit Span	19.3	4.0	0.78	0.84	0.97	0.90	0.83
Coding	30.3	6.5	0.63	0.50	0.37	0.33	0.18
Vocabulary	18.0	4.8	0.66	0.69	0.47	0.78	0.36
Word Reasoning	16.5	3.2	0.53	0.56	0.43	0.57	0.27
Picture Completion	12.4	2.9	0.65	0.77	0.58	0.80	0.62
Picture Concepts	22.9	5.1	0.61	0.85	0.65	0.79	0.83
Symbol Search	18.7	2.6	0.30	0.21	0.17	0.24	0.19
Information	10.8	2.3	0.41	0.35	0.37	0.36	0.24
Cancellation	51.4	3.7	0.10	0.06	0.06	0.07	0.05
Comprehension	12.2	3.6	0.61	0.54	0.32	0.54	0.40
Arithmetic	14.6	2.7	0.67	0.57	0.60	0.51	0.46

Source: Estimates using Chilean Wechsler Intelligence Scale for Children (WISC-V) survey.
Note: Test scores are assumed to be arbitrary scaled measures of skills θ_{it}. The test score M_{itm}^A of child i at grade t in test m follows the structure $M_{itm}^A = \mu_m^A + \lambda_m^A \log \theta_{it} + \varepsilon_{itm}$, where ε_{itm} is measurement error and μ_m^A and λ_m^A are the location and scale parameters. The estimates of location and scale are for the initial period; when children are 8 years old. The signal share (out of the measure variance) is $1 - \text{Var}(\varepsilon_{itm})/\text{Var}(M_{itm}^A)$. See the measurement equation (1.5.1) for additional details. Back to Section 1.7.1.

Table S.1.4:
Measurement system of skills: Administrative data

	Second grade			Fourth grade			Sixth grade		
	μ_{mt}	λ_{mt}	Signal share	μ_{mt}	λ_{mt}	Signal share	μ_{mt}	λ_{mt}	Signal share
Math SIMCE	-	-	-	217.6	32.9	0.64	163.5	35.5	0.65
Language SIMCE	253.1	29.2	0.35	222.3	34.4	0.63	155.8	37.1	0.66
Natural sciences SIMCE	-	-	-	-	-	-	162.4	36.2	0.60
Social sciences SIMCE	-	-	-	216.1	30.5	0.66	162.0	35.5	0.64
Math. grade	5.6	0.7	0.69	4.8	0.5	0.55	3.9	0.5	0.47
Language grade	5.5	0.8	0.98	4.8	0.5	0.62	4.1	0.5	0.50
	Eighth grade			Tenth grade					
	μ_{mt}	λ_{mt}	Signal share	μ_{mt}	λ_{mt}	Signal share			
Math SIMCE	165.3	32.0	0.65	41.4	56.0	0.66			
Language SIMCE	146.8	32.9	0.62	85.2	41.2	0.55			
Natural sciences SIMCE	176.4	29.7	0.58	75.5	41.9	0.64			
Social sciences SIMCE	170.6	30.2	0.54	97.7	38.3	0.53			
Math grade	3.8	0.4	0.36	2.5	0.6	0.35			
Language grade	4.0	0.4	0.41	3.1	0.5	0.42			

Source: Estimates using SIMCE and SIGE administrative databases.
Note: Test scores are assumed to be arbitrary scaled measures of skills θ_{it}. The test score M_{itm} of student i at grade t in subject m follows the structure $M_{itm} = \mu_{tm} + \lambda_{tm} \log \theta_{it} + \varepsilon_{itm}$, where ε_{itm} is measurement error and μ_{mt} and λ_{mt} are the location and scale parameters. The signal share (out of the measure variance) is $1 - \text{Var}(\varepsilon_{itm})/\text{Var}(M_{itm})$. See the measurement equation (1.5.4) for additional details. Back to Section 1.7.1.

Table S.1.5:
Skill formation technology

	School grade			
	Fourth (1)	Sixth (2)	Eighth (3)	Tenth (4)
$\log \theta_{it}$ (Skill)	0.795	0.495	0.588	0.493
	(0.017)	(0.008)	(0.011)	(0.012)
	[0.761;0.829]	[0.480;0.511]	[0.569;0.61]	[0.471;0.519]
$\log^2 \theta_{it}$	0.066	0.063	0.034	0.003
	(0.003)	(0.001)	(0.002)	(0.002)
	[0.06;0.071]	[0.06;0.066]	[0.031;0.038]	[0.000;0.007]
h_{it} (parental time - daily hours)	0.442	0.104	0.278	0.144
	(0.051)	(0.020)	(0.037)	(0.031)
	[0.340;0.547]	[0.069;0.144]	[0.203;0.349]	[0.082;0.206]
h_{it}^2	-0.056	-0.035	-0.059	-0.048
	(0.009)	(0.005)	(0.009)	(0.008)
	[-0.074;-0.039]	[-0.045;-0.025]	[-0.077;-0.040]	[-0.064;-0.032]
e_{it} (child time - daily hours)	-1.787	0.462	0.096	-0.095
	(1.074)	(0.043)	(0.043)	(0.046)
	[-3.968;0.361]	[0.375;0.544]	[0.003;0.18]	[-0.183;-0.006]
e_{it}^2	8.106	-0.153	-0.132	0.041
	(3.103)	(0.017)	(0.029)	(0.019)
	[1.968;14.726]	[-0.185;-0.117]	[-0.187;-0.077]	[0.001;0.075]
$\log C_{it}$ (classroom effects)	0.577	0.479	0.541	0.422
	(0.017)	(0.008)	(0.012)	(0.012)
	[0.542;0.608]	[0.461;0.494]	[0.516;0.565]	[0.394;0.444]

(*table continues in next page*)

Table S.1.5: — *Continued*
Skill formation technology

	School grade			
	Fourth (1)	Sixth (2)	Eighth (3)	Tenth (4)
$\log\theta_{it} \times \log C_{it}$	-0.042	0.009	0.007	0.002
	(0.003)	(0.003)	(0.003)	(0.003)
	[-0.048;-0.036]	[0.004;0.013]	[0.000;0.014]	[-0.003;0.009]
$h_{it} \times \log C_{it}$	0.014	-0.003	0.013	0.033
	(0.007)	(0.004)	(0.003)	(0.010)
	[0.004;0.028]	[-0.011;0.006]	[0.007;0.020]	[0.015;0.052]
$e_{it} \times \log C_{it}$	-0.130	0.024	-0.036	-0.031
	(0.083)	(0.009)	(0.007)	(0.007)
	[-0.292;0.040]	[0.008;0.042]	[-0.052;-0.022]	[-0.046;-0.018]
$h_{it} \times e_{it}$	-0.610	-0.030	0.052	0.092
	(0.176)	(0.026)	(0.013)	(0.022)
	[-0.965;-0.248]	[-0.081;0.025]	[0.026;0.077]	[0.049;0.142]
$h_{it} \times \log\theta_{it}$	-0.013	0.061	-0.021	-0.009
	(0.005)	(0.005)	(0.003)	(0.005)
	[-0.023;-0.002]	[0.049;0.070]	[-0.028;-0.015]	[-0.019;0.002]
$e_{it} \times \log\theta_{it}$	0.332	0.062	0.103	0.056
	(0.094)	(0.007)	(0.007)	(0.007)
	[0.171;0.529]	[0.048;0.077]	[0.090;0.117]	[0.043;0.069]
N	407,720	596,617	457,782	336,470

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. Sample consists of students in classrooms with at least 10 students with non-missing values in test scores and time investment questions. All specifications include an intercept, student's gender and age, mother answered questionnaire indicator, parents' education and age, and indicator variables for missing controls. Back to Section 1.7.2.

Table S.1.6:
Skill formation technology - specification with within classroom inputs

	School grade			
	Fourth	Sixth	Eighth	Tenth
	(1)	(2)	(3)	(4)
$\log\theta_{it}$ (Skill)	0.815	0.487	0.497	0.510
	(0.023)	(0.009)	(0.033)	(0.035)
	[0.765;0.858]	[0.467;0.504]	[0.476;0.508]	[0.485;0.533]
$\log^2\theta_{it}$	0.062	0.059	0.068	0.000
	(0.005)	(0.002)	(0.005)	(0.002)
	[0.053;0.071]	[0.055;0.062]	[0.065;0.073]	[-0.003;0.003]
h_{it} (parental time -	0.073	0.119	0.035	0.009
daily hours)	(0.053)	(0.020)	(0.033)	(0.027)
	[-0.018;0.181]	[0.073;0.152]	[-0.027;0.100]	[-0.06;0.044]
h^2_{it}	-0.009	-0.030	-0.016	0.002
	(0.011)	(0.004)	(0.008)	(0.006)
	[-0.029;0.014]	[-0.037;-0.021]	[-0.031;0]	[-0.009;0.013]
e_{it} (child time -	2.200	0.576	0.238	0.045
daily hours)	(1.427)	(0.041)	(0.059)	(0.042)
	[-0.507;4.984]	[0.516;0.682]	[0.110;0.346]	[-0.049;0.114]
e^2_{it}	-5.263	-0.204	-0.100	0.038
	(4.277)	(0.020)	(0.043)	(0.020)
	[-13.523;2.677]	[-0.254;-0.174]	[-0.187;-0.013]	[0.004;0.086]
$\log T_{it}$ (teachers effects)	0.437	0.375	0.493	0.350
	(0.023)	(0.012)	(0.039)	(0.029)
	[0.388;0.476]	[0.345;0.391]	[0.456;0.545]	[0.313;0.385]

(*table continues in next page*)

Table S.1.6: — *Continued*
Skill formation technology - specification with within classroom inputs

	School grade			
	Fourth (1)	Sixth (2)	Eighth (3)	Tenth (4)
$\log\theta_{it}\times\log T_{it}$	-0.040	0.008	-0.037	-0.019
	(0.005)	(0.003)	(0.006)	(0.004)
	[-0.050;-0.031]	[0.001;0.013]	[-0.052;-0.031]	[-0.026;-0.011]
$h_{it}\times\log T_{it}$	-0.006	-0.023	0.001	0.000
	(0.007)	(0.007)	(0.004)	(0.006)
	[-0.021;0.007]	[-0.038;-0.008]	[-0.008;0.008]	[-0.013;0.011]
$e_{it}\times\log T_{it}$	0.165	0.061	0.006	0.000
	(0.119)	(0.011)	(0.010)	(0.008)
	[-0.048;0.436]	[0.039;0.086]	[-0.016;0.023]	[-0.014;0.018]
$h_{it}\times e_{it}$	0.151	-0.067	0.070	-0.029
	(0.233)	(0.024)	(0.021)	(0.021)
	[-0.271;0.636]	[-0.108;-0.019]	[0.024;0.106]	[-0.062;0.020]
$h_{it}\times\log\theta_{it}$	-0.025	0.063	-0.007	0.009
	(0.007)	(0.007)	(0.003)	(0.005)
	[-0.037;-0.010]	[0.049;0.075]	[-0.013;0.001]	[0.001;0.022]
$e_{it}\times\log\theta_{it}$	0.267	0.070	0.055	0.050
	(0.133)	(0.008)	(0.011)	(0.008)
	[0.007;0.525]	[0.055;0.087]	[0.034;0.077]	[0.032;0.060]
N	199,000	452,575	328,888	253,721

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. Sample consists of students in classrooms with at least 10 students with non-missing values in test scores and time investment questions and their math teacher is observed teaching at least two classrooms in different calendar years. All specifications include an intercept, student's gender and age, mother answered questionnaire indicator, parents' education and age, math teacher teaching experience and tenure at school (second order polynomial), class size, class share male students, poor students and parents education (less than high school, high school, and more than high school), class average parents age, household income, peers' skills in previous grade, and parental and student time, number of teachers and subjects and indicator variables for missing controls. Back to Section 1.7.2.

Table S.1.7:
Classroom and teacher value added

	School grade							
	Fourth		Sixth		Eighth		Tenth	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
Classroom VA	0.39	0.26	0.31	0.26	0.32	0.28	0.37	0.33
Classrooms	31,639	31,499	38,102	38,438	30,771	30,579	29,477	29,199
Students	721,942	719,470	876,631	882,752	675,334	670,766	701,621	693,673
Teacher VA								
general	0.36	0.21	0.28	0.19	0.30	0.22	0.33	0.24
within school	0.18	0.13	0.22	0.17	0.19	0.20	0.24	0.33
Teachers	3,362	3,174	5,268	4,674	3,517	3,159	4,514	4,607
Classrooms	9,153	8,566	19,229	16,413	11,509	10,384	22,557	22,769
Schools	1,478	1,380	2,706	2,324	1,844	1,637	1,800	1,850
Students	238,681	224,218	462,374	403,760	270,196	246,726	522,911	517,716

Source: SIMCE and SIGE administrative data.

Note: Table reports classroom and teacher value-added (VA), estimated as the dispersion of classroom or teacher fixed effects in the regression of test scores on second order polynomials of previous scores (both math and language), parents' education and age, mom answered survey indicator, and household income. In the case of teacher value-added, the specification includes classroom average of parents' education, age, household income, peers' previous test score, share poor classmates, class size, second order polynomial teacher's tenure at school and teaching experience, and teacher-student gender match indicator. Additionally, specification of teacher VA within school includes school fixed effects. Classroom and teacher VA is adjusted for measurement error by subtracting the mean error variance (the average of the squared standard errors on the estimated fixed effects) from the variance of the fixed effects. Fixed effects' standard errors are estimated by bootstrap. Sample classroom value-added includes all students in classrooms with at least 10 students. Sample teacher value-added includes all students in classrooms with at least 10 students, and their schools have enrollment all years in the period, and at least two teachers for whom there are observation in at least two different points in their careers. Back to Section 1.7.2.

Table S.1.8:
Parental and child responses to classroom and teacher effects

	$\log E_{it} = \log C_{it}$ (classroom effects)		$\log E_{it} = \log T_{it}$ (teachers effects)	
	Parental time (weekly hours)	Child time (weekly hours)	Parental time (weekly hours)	Child time (weekly hours)
	(1)	(2)	(3)	(4)
$\log E_{it}$	-1.353	0.240	-0.557	0.172
	(0.129)	(0.081)	(0.124)	(0.054)
	[-1.556;-1.143]	[0.170;0.306]	[-0.753;-0.293]	[0.101;0.282]
$\log E_{it} \times 1\{\text{Grade} = 6\}$	0.723	0.137	-0.073	0.091
	(0.112)	(0.043)	(0.107)	(0.036)
	[0.508;0.942]	[0.058;0.182]	[-0.290;0.136]	[-0.018;0.125]
$\log E_{it} \times 1\{\text{Grade} = 8\}$	1.112	0.049	0.473	0.055
	(0.109)	(0.046)	(0.117)	(0.046)
	[0.952;1.370]	[-0.030;0.155]	[0.249;0.698]	[-0.028;0.146]
$\log E_{it} \times 1\{\text{Grade} = 10\}$	1.932	0.168	0.806	-0.269
	(0.148)	(0.083)	(0.142)	(0.063)
	[1.740;2.215]	[0.127;0.266]	[0.482;1.008]	[-0.418;-0.185]
N	1,079,935	1,079,935	555,415	555,415

Note: Schools networks-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets, respectively. Sample consists of students in classrooms with at least 10 students with non-missing values in test scores and time investment questions and that at least the student is observed in two grades between fourth and tenth grade. All specifications include an intercept, current skill interacted with school grade, household income (second order polynomial), indicator for missing controls and school grade and student fixed effects. The specifications of columns (3) and (4) also include math teacher's teaching experience and tenure at school (second order polynomial), class size, class share of male students, poor students and parents education (less than high school, high school, and more than high school), class average parents' age, household income, peers' skills in previous grade and parental and student time, number of teachers and subjects and indicator variables for missing controls. Back to Section 1.7.3.

Table S.2.9:
Standard deviation of the skill technology shock

	School grades			
	4th grade	6th grade	8th grade	10th grade
$\sigma_{\nu,t}$	1.209	0.869	0.982	0.711
	(0.005)	(0.003)	(0.002)	(0.005)
	[1.202;1.222]	[0.863;0.874]	[0.977;0.986]	[0.702;0.722]

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. The standard deviation of the skill technology shock is estimated as the standard deviation of the residuals from the specification in Table S.1.5. Back to Section 2.2.3.

Table S.2.10:
Household income process

	School grades			
	4th grade	6th grade	8th grade	10th grade
$\overline{y}_t$	-	3.434	3.365	3.383
		(0.026)	(0.024)	(0.025)
		[3.397;3.498]	[3.336;3.43]	[3.335;3.433]
ρ_t^y	-	0.743	0.749	0.748
		(0.002)	(0.002)	(0.002)
		[0.738;0.746]	[0.744;0.751]	[0.744;0.751]
$\sigma^2_{\omega,t}$	-	0.269	0.268	0.285
		(0.002)	(0.002)	(0.002)
		[0.266;0.273]	[0.266;0.273]	[0.282;0.289]

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parenthesis and brackets respectively. The household income process is $\log y_{it} = \overline{y}_t + \rho_t \log y_{it} + \omega_{it}$, where $\omega_{it} \sim N(0, \sigma^2_{\omega,t})$. It is estimated with OLS estimator. Household income is measured as monthly income in Chilean pesos. For details of the household income process see equation (2.2.5). Back to Section 2.2.3.

Table S.2.11:
Classroom process

	School grades			
	4th grade	6th grade	8th grade	10th grade
Constant	-0.260	-0.525	-0.610	-0.804
	(0.0270)	(0.0213)	(0.0272)	(0.0295)
	[-0.315;-0.208]	[-0.57;-0.485]	[-0.664;-0.56]	[-0.866;-0.752]
Montly HH income (ths. CLP)	0.250	0.258	0.355	0.333
	(0.0185)	(0.0145)	(0.0140)	(0.0116)
	[0.216;0.288]	[0.229;0.285]	[0.33;0.383]	[0.312;0.357]
Father HS	0.019	0.084	0.099	0.137
	(0.0083)	(0.0078)	(0.0088)	(0.0091)
	[0.003;0.035]	[0.069;0.1]	[0.083;0.115]	[0.12;0.156]
Father >HS	0.029	0.176	0.177	0.230
	(0.0120)	(0.0102)	(0.0120)	(0.0129)
	[0.006;0.053]	[0.157;0.197]	[0.153;0.201]	[0.206;0.256]
Mother HS	0.055	0.143	0.161	0.202
	(0.0088)	(0.0075)	(0.0079)	(0.0091)
	[0.038;0.073]	[0.127;0.157]	[0.146;0.176]	[0.185;0.221]
Mother >HS	0.066	0.217	0.239	0.304
	(0.0114)	(0.0101)	(0.0106)	(0.0123)
	[0.043;0.089]	[0.197;0.237]	[0.218;0.26]	[0.282;0.331]
Age parent	0.002	0.005	0.006	0.008
	(0.0004)	(0.0003)	(0.0004)	(0.0004)
	[0.002;0.003]	[0.004;0.005]	[0.005;0.007]	[0.008;0.009]
σ^2_Δ	0.926	0.901	0.875	0.864
	(0.0073)	(0.0068)	(0.0097)	(0.0094)
	[0.909;0.939]	[0.887;0.912]	[0.855;0.893]	[0.843;0.881]
N	407,720	596,617	457,782	336,470

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. The classroom process is $\log C_{it} = \kappa_t' \mathbf{x}_{it} + \Delta_{it}$, where $\Delta_{it} \sim N(0, \sigma^2_{\Delta,t})$, is estimated with OLS estimator. For additional details see equation (2.2.4). HS refers to high school education. Monthly HH income (ths. CLP) refers to monthly household income in thousands of 2018 Chilean pesos. Back to Section 2.2.3.

Table S.2.12:
Estimates of preference parameters

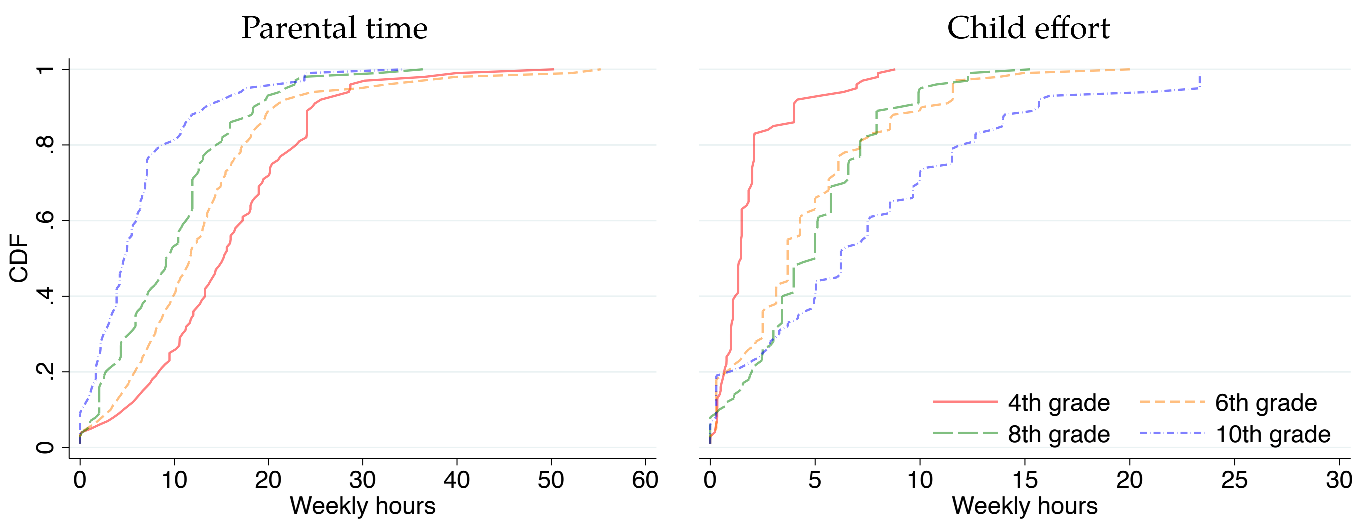
	School grade							
	Fourth		Sixth		Eighth		Tenth	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
CRRA utility parameter on skills:								
ϕ_{1t}	0.566	(0.0001)	0.053	(0.0001)	0.383	(0)	0.061	(0.0108)
Parent time disutility costs parameters: $\tilde{\phi}_{2it}$								
Constant	4.579	(0.1081)	4.610	(0.0145)	2.336	(0.0549)	5.859	(0.0698)
HH income	-1.798	(0.0048)	-0.040	(0.0052)	-0.261	(0.0004)	-0.023	(0.0003)
Father HS	-2.228	(0.0079)	-0.279	(0.0002)	-0.003	(0.0061)	-1.781	(0.007)
Father >HS	0.000	(0.0002)	-1.451	(0.0313)	-0.011	(0.0003)	-0.004	(0.0001)
Mother HS	-0.046	(0.0001)	-2.803	(0.0074)	-0.062	(0.0009)	-0.391	(0.0012)
Mother >HS	-0.412	(0.0005)	-0.007	(0.0007)	-0.001	(0.0003)	-0.371	(0.0001)
Parent age	-1.414	(0.0041)	-2.218	(0.0006)	-2.215	(0.0065)	-1.332	(0.0036)
Student time disutility costs parameters: $\tilde{\phi}_{3it}$								
Constant	5.936	(0.1383)	4.285	(0.0144)	4.913	(0.0112)	5.257	(0.1022)
HH income	1.779	(0.0035)	1.413	(0.0050)	1.934	(0.0028)	-2.550	(0.0057)
Father HS	1.765	(0.0020)	-1.495	(0.0015)	0.544	(0.0023)	0.620	(0.0045)
Father >HS	1.373	(0.0016)	-1.535	(0.0041)	0.356	(0.0027)	1.987	(0.0035)
Mother HS	1.985	(0.0020)	-0.237	(0.0067)	-2.509	(0.0049)	1.114	(0.0038)
Mother >HS	1.478	(0.0016)	-0.236	(0.0019)	-1.699	(0.0038)	1.762	(0.0037)
Parent age	1.991	(0.0047)	-0.569	(0.0002)	-1.323	(0.0023)	-2.230	(0.0055)

	Estimate	SE
Terminal period parameter: ϕ_4	1.007	(0.0015)
Variance υ_i: σ^2_υ	0.025	(0.0001)
Variance ι_i: σ^2_ι	0.059	(0.0020)

Note: Schools network-clustered bootstrapped standard errors in parentheses. The preference parameters are specified as: CRRA parameter on skills ϕ_{1t}, parental and student time disutility costs $\phi_{i2t} = \exp(\tilde{\phi}'_{2t}\mathbf{x}_{it} + \upsilon_i)$ and $\phi_{i3t} = \exp(\tilde{\phi}'_{3t}\mathbf{x}_{it} + \iota_i)$, respectively, where $\upsilon_i \sim N(0, \sigma^2_\upsilon)$ and $\iota_i \sim N(0, \sigma^2_\iota)$. Terminal period parameter ϕ_{4t}. For additional detail see Section 2.2.1. The vector $\mathbf{x}_{it}$ includes a one, household income, parents' education and age. HS refers to high school education. HH income refers to monthly household income measured in thousands of Chilean pesos. Back to Section 2.2.3.

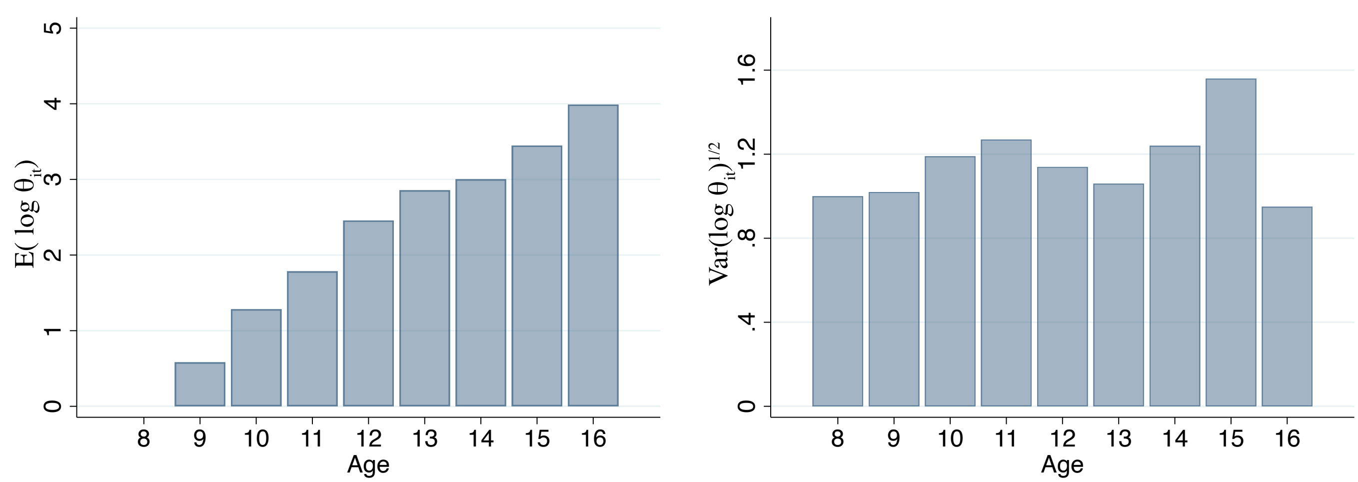
Supplemental Figures

Figure S.1.1:
Parental and child time investment distributions



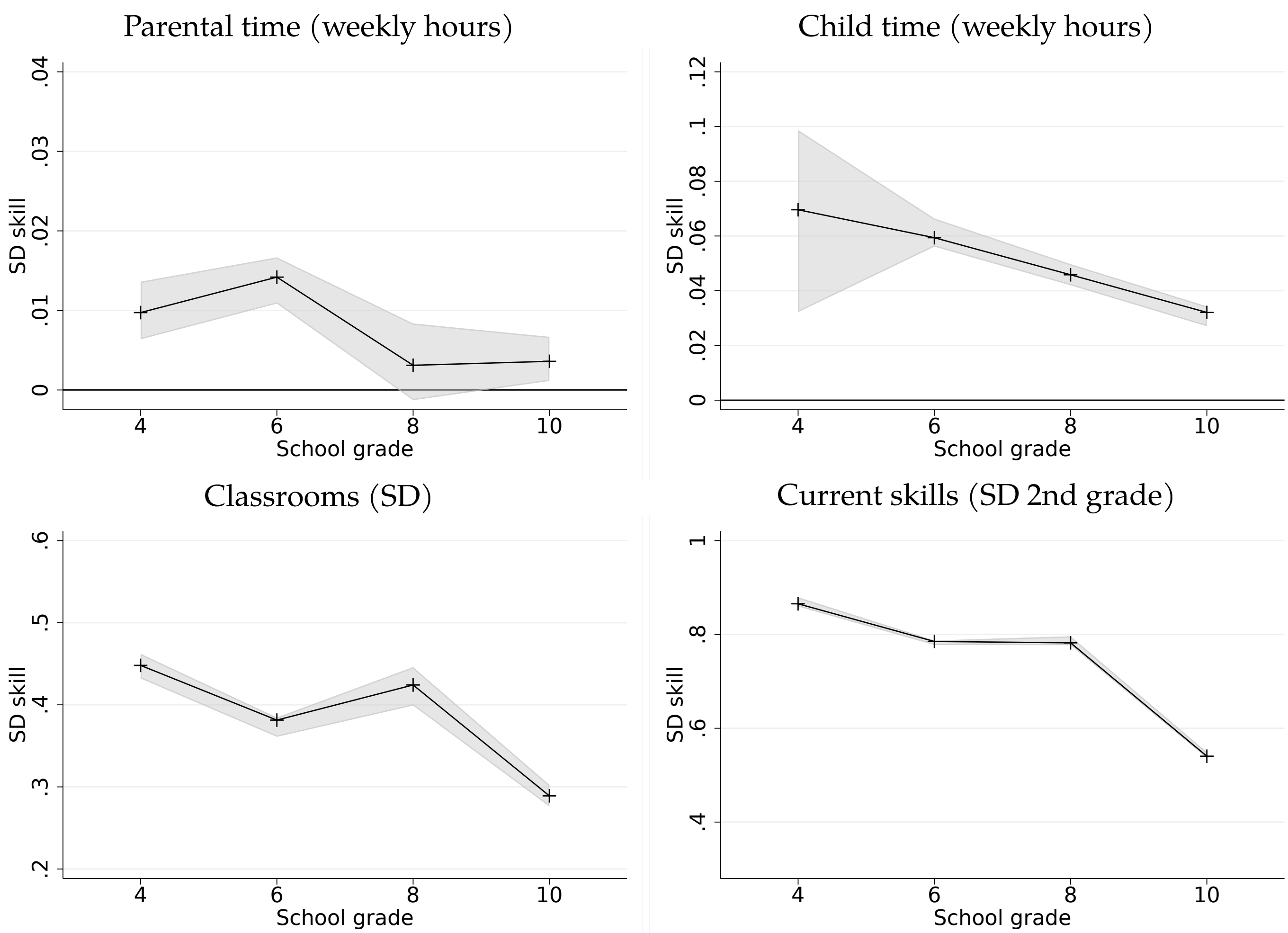
Note: Calculations are based on the Chilean Time Use Survey 2015. The survey reports hours spend in activities in the last week and weekend day. Parental time refers to hours parents spent in activities with their children and child time is hours a child spends studying, doing homework or other academic activities outside school. I transformed the reported time to hours per week by multiplying by 5 and 2 times during the week and weekend day, respectively. Back to Section 1.3.2.

Figure S.1.2:
Expected value and standard deviation of cognitive skills



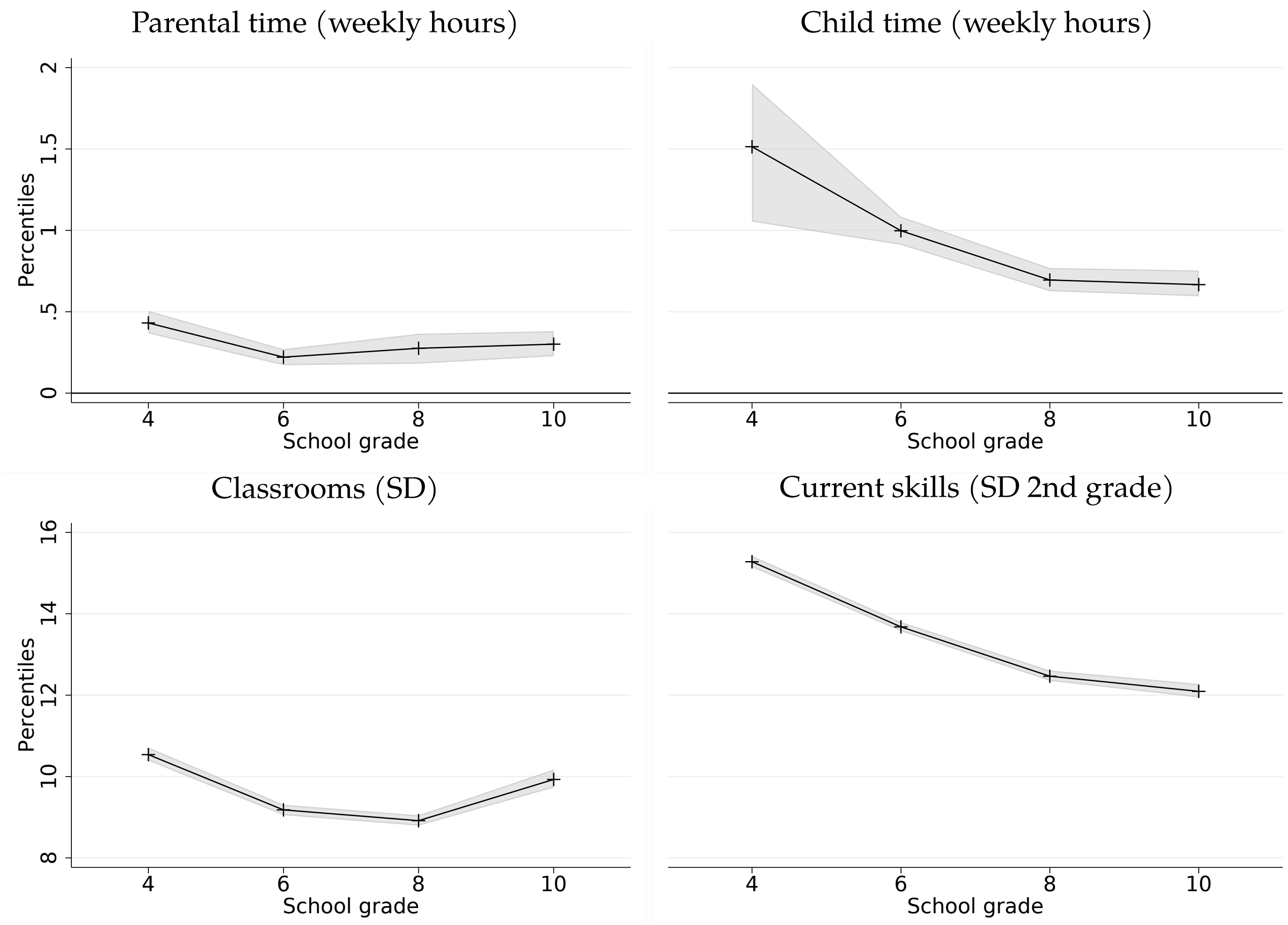
Note: Estimated with the Chilean WISC-V cognitive development test survey. The estimation consist on replacing the moments of equation (1.5.3) with their sample analogs and the measurement system's parameters presented in Table S.1.3. Back to Section 1.7.1.

Figure S.1.3:
Sample average marginal effects of skill inputs (SD)
specification with within classroom components



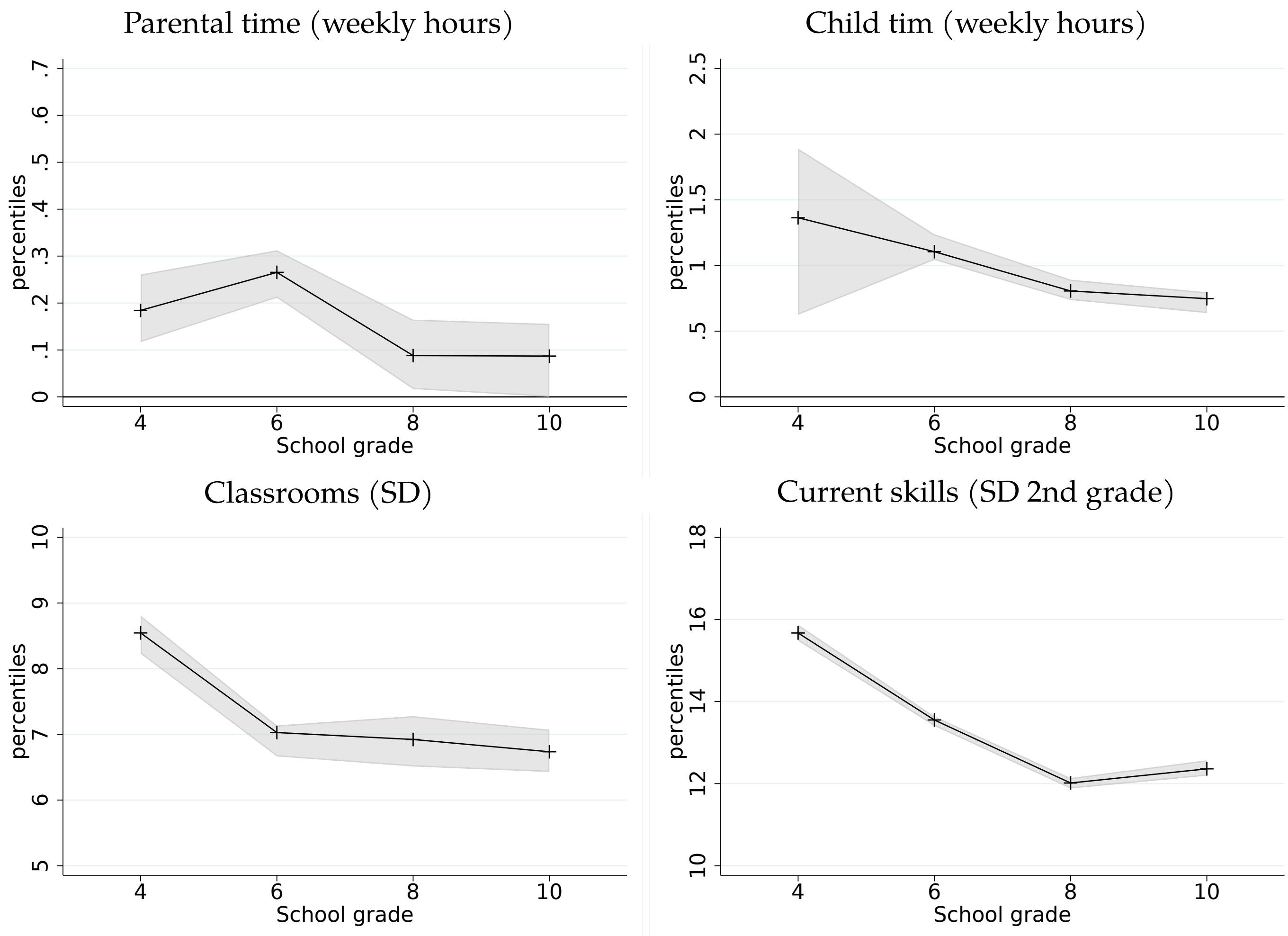
Note: The values on these graphs show the average marginal effect calculated using the estimates from the specifications in Table S.1.6. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student's marginal effect using each input's analogous equation (1.7.14) and calculate the average over the sample. Back to Section 1.7.2.

Figure S.1.4:
Sample average marginal effects of skill inputs (Percentiles)



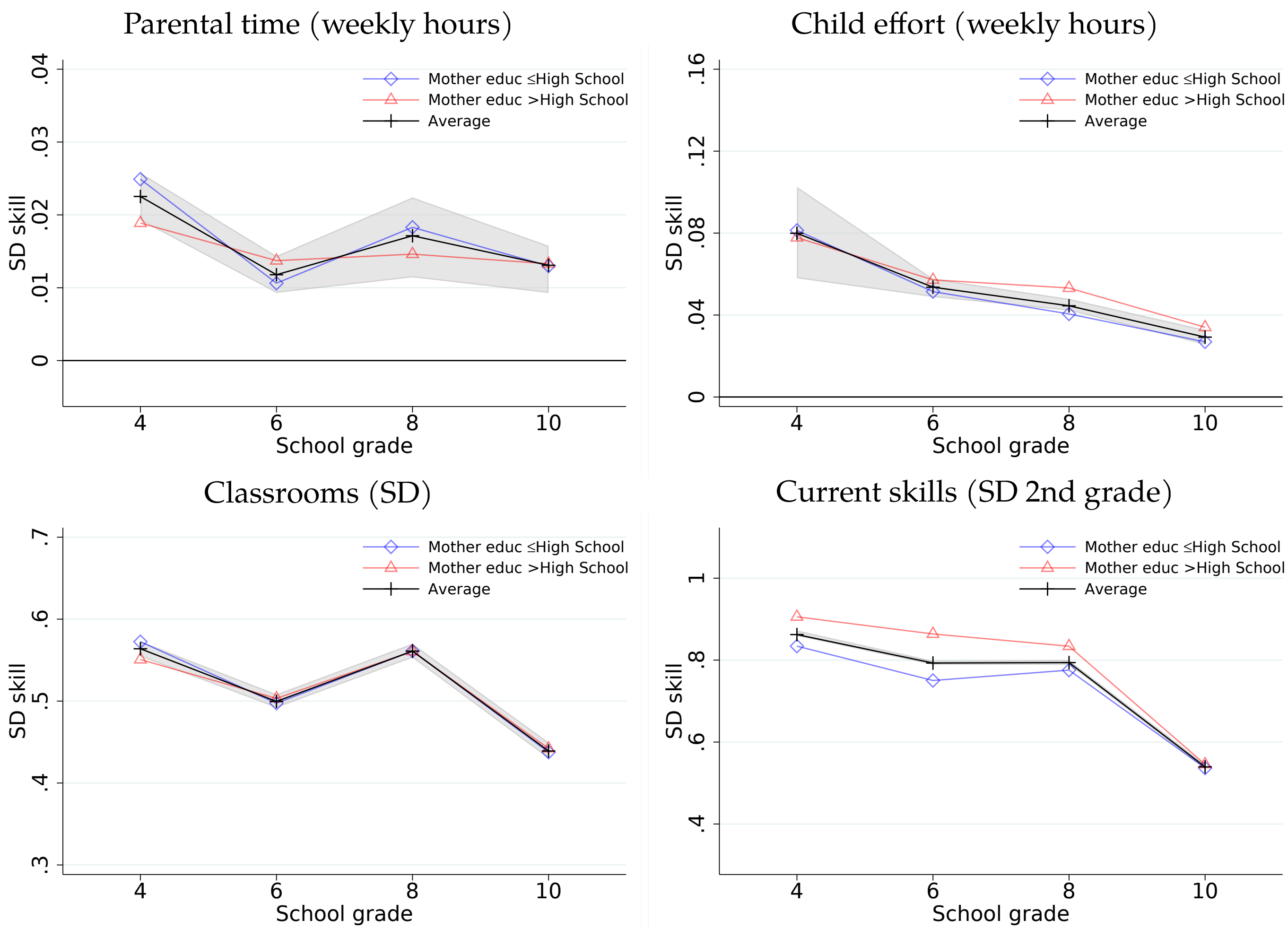
Note: The values on these graphs show the average shift in percentiles across the skill distribution of an additional unit of input. Calculated using the estimates from the specifications in Table S.1.5. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student's percentile shift implied by the marginal effect from the each input's analogous equation (1.7.14) and calculate the average over the sample. Back to Section 1.7.2.

Figure S.1.5:
Sample average marginal effects of skill inputs (Percentiles)
specification with within classroom components



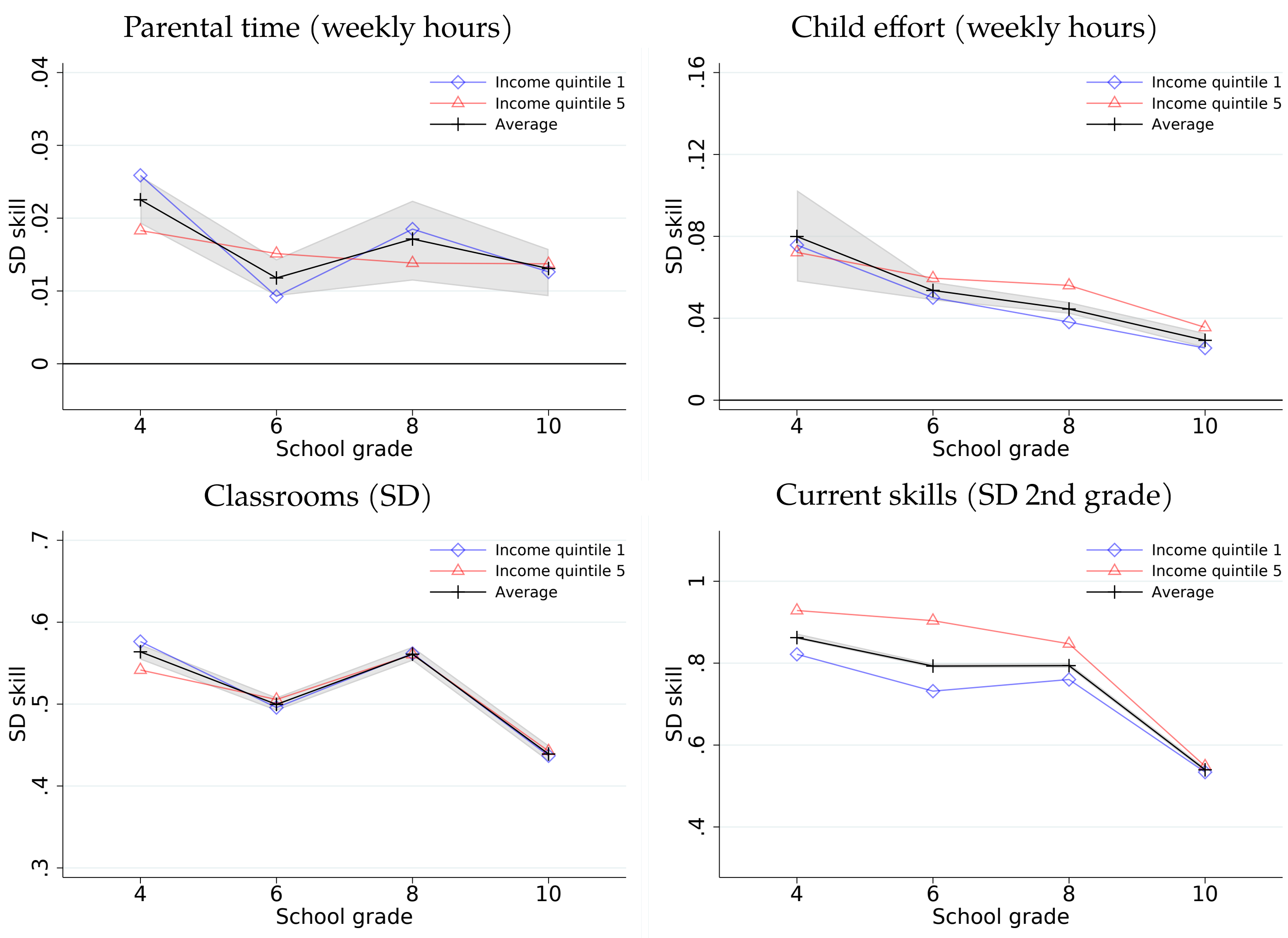
Note: The values on these graphs show the average shift in percentiles across the skill distribution of an additional unit of input. Calculated using the estimates from the specifications in Table S.1.6. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student's percentile shift implied by the marginal effect from the each input's analogous equation (1.7.14) and calculate the average over the sample. Back to Section 1.7.2.

Figure S.1.6:
Sample average marginal effects of skill formation inputs (in SD log skill 2nd grade) by mother education



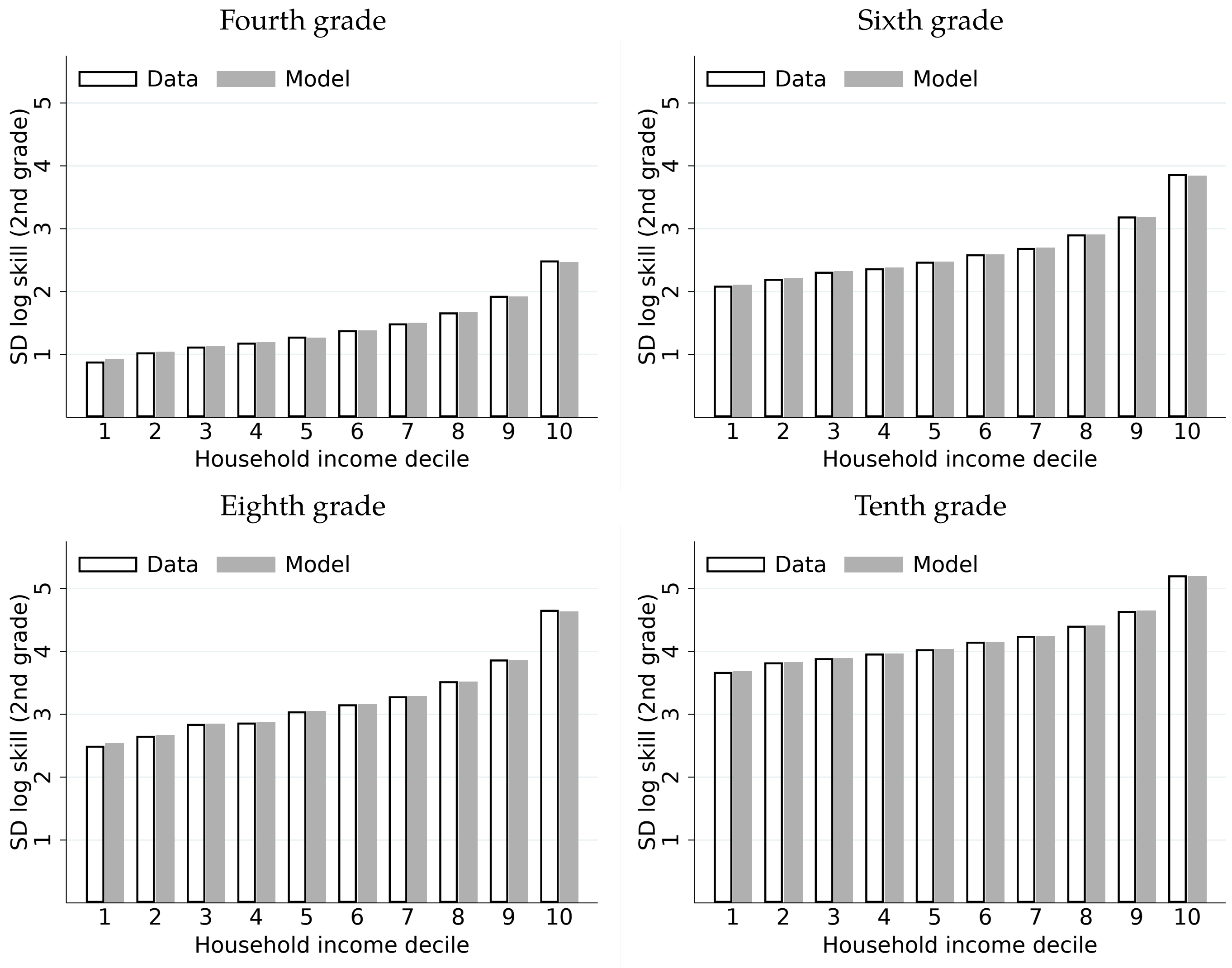
Note: The values on these graphs present the average marginal effect calculated using the estimates from the specifications in Table S.1.5. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student's marginal effect using each input analogous equation (1.7.14) and calculate the average over the sample for the grand average (black + line). And for the blue diamond and red triangles, I calculate the average within each mother with high school or less and more than high school education, respectively.

Figure S.1.7:
Sample average marginal effects of skill formation inputs (in SD log skill 2nd grade) by household income



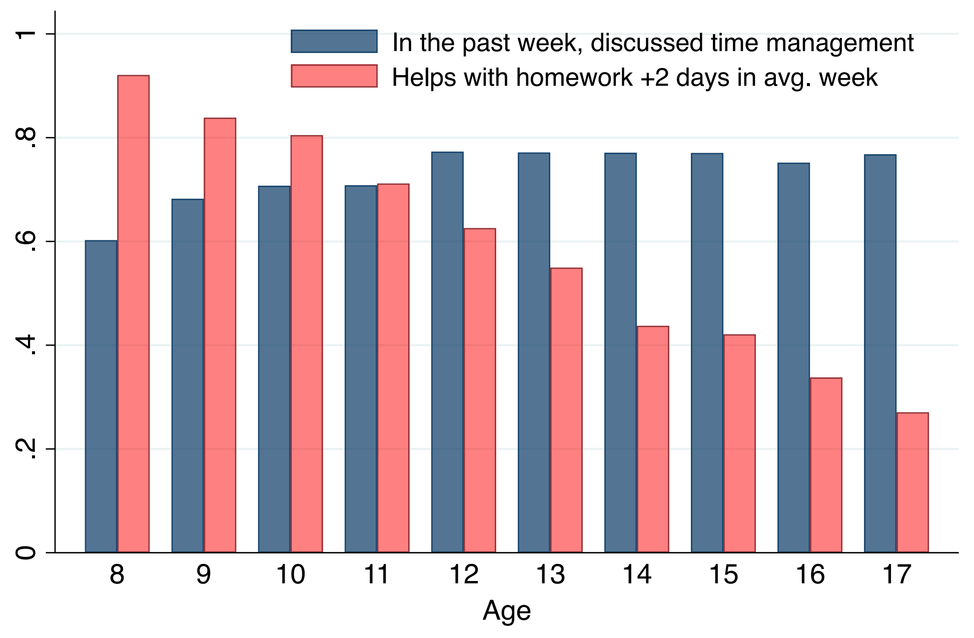
Note: The values on these graphs present the average marginal effect calculated using the estimates from the specifications in Table S.1.5. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student's marginal effect using each input analogous equation (1.7.14) and calculate the average over the sample for the grand average (black + line). And for the blue diamond and red triangles, I calculate the average within the first and fifth household income quintile, respectively.

Figure S.1.8:
Validation test: skill formation technology
Out-of-sample fit of skills by household income



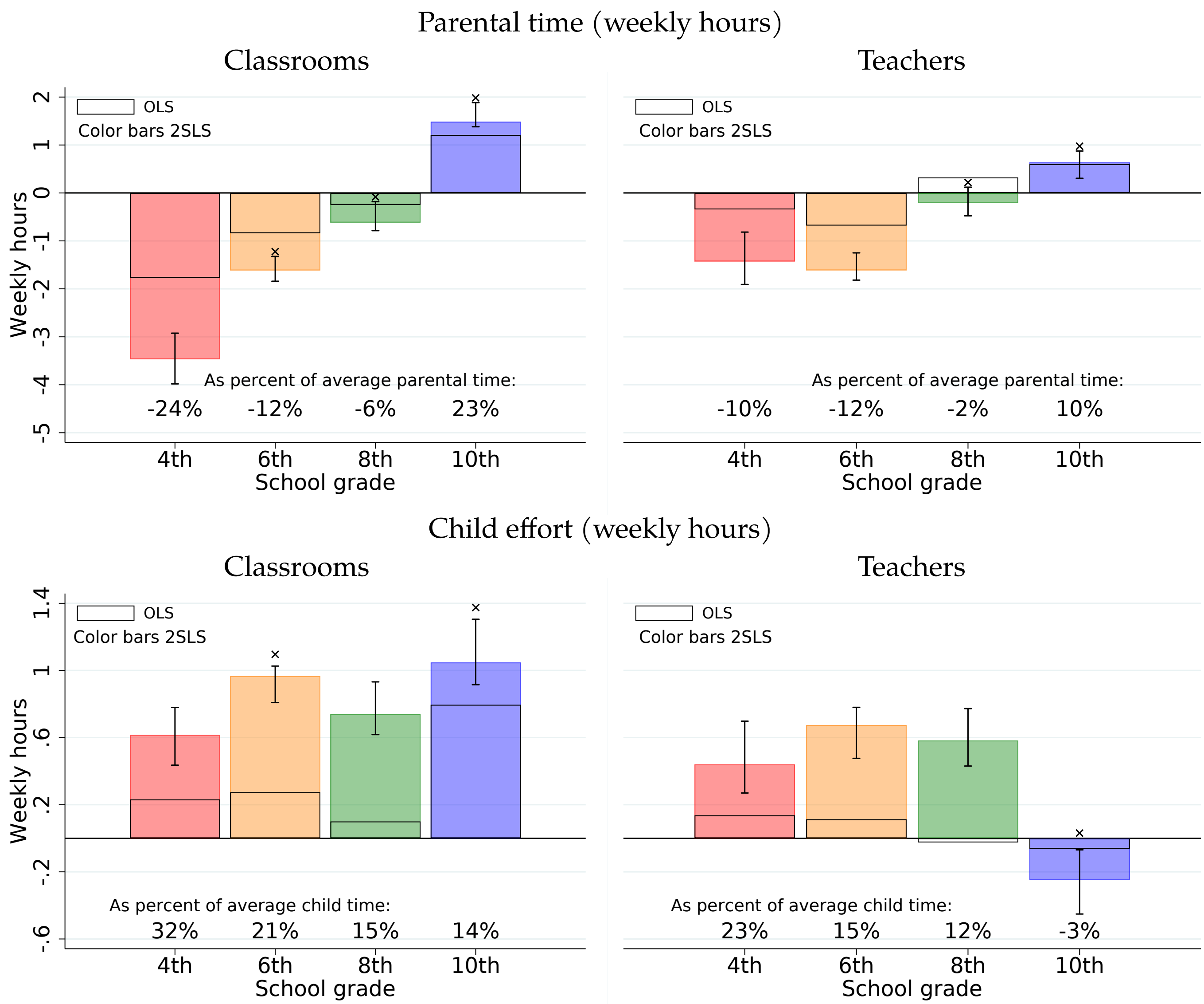
Note: The graphs compares the average predicted skill by the estimated skill technology (Model) and the average of the skill estimated in the data (Data) by household income deciles. Back to Section 1.7.2.

Figure S.1.9:
Activities of parental time across age of children



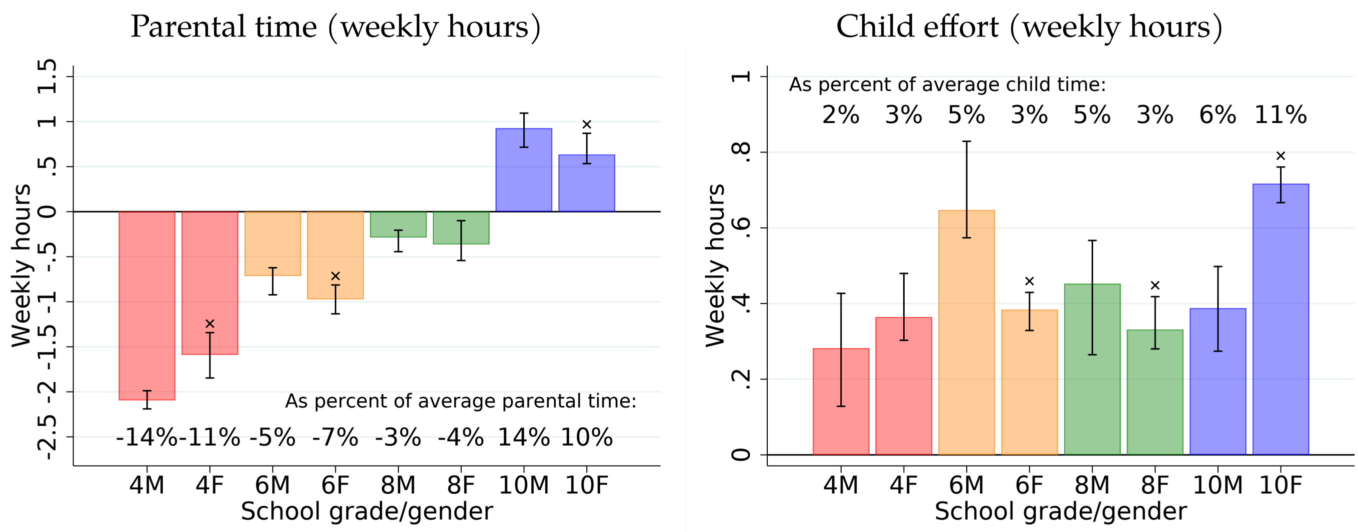
Source: National Household Education Survey (NHES), 2016.
Note: This figure shows share of parents that in the previous week discussed time management with their children and the share of parents (or other caretaker) that help the child with her/his homework more than 2 days in an average week. Back to Section 1.7.3.

Figure S.1.10:
Parental time and child effort responses
to reassignment from 10th to 90th percentile of classroom and teacher distributions



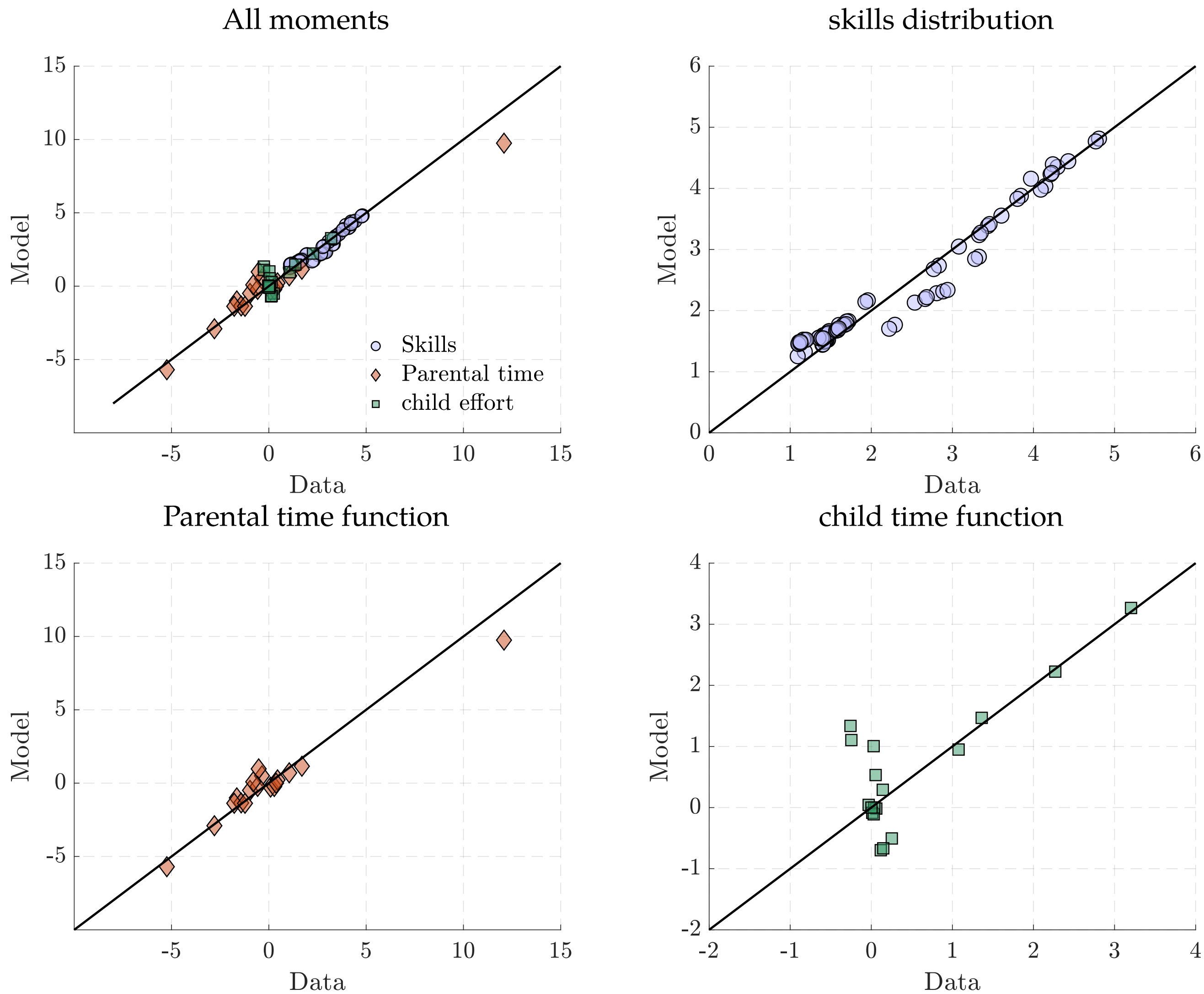
Note: The values on the graphs are calculated using the estimates of Table S.1.8. The top and bottom panels reports parental and student responses, respectively. The vertical lines are school network-clustered bootstrapped 95% confidence intervals. The symbol on top (x) indicates that the difference between the response with that of fourth grade is statistically significant at 1%. The values on the bottom of each plot are the response as a percent of the average time investment at each grade. Back to Section 1.7.3.

Figure S.1.11:
Household responses by child gender (F=female, M=male)
to reassignment from 25th to 75th percentile of classroom distributions



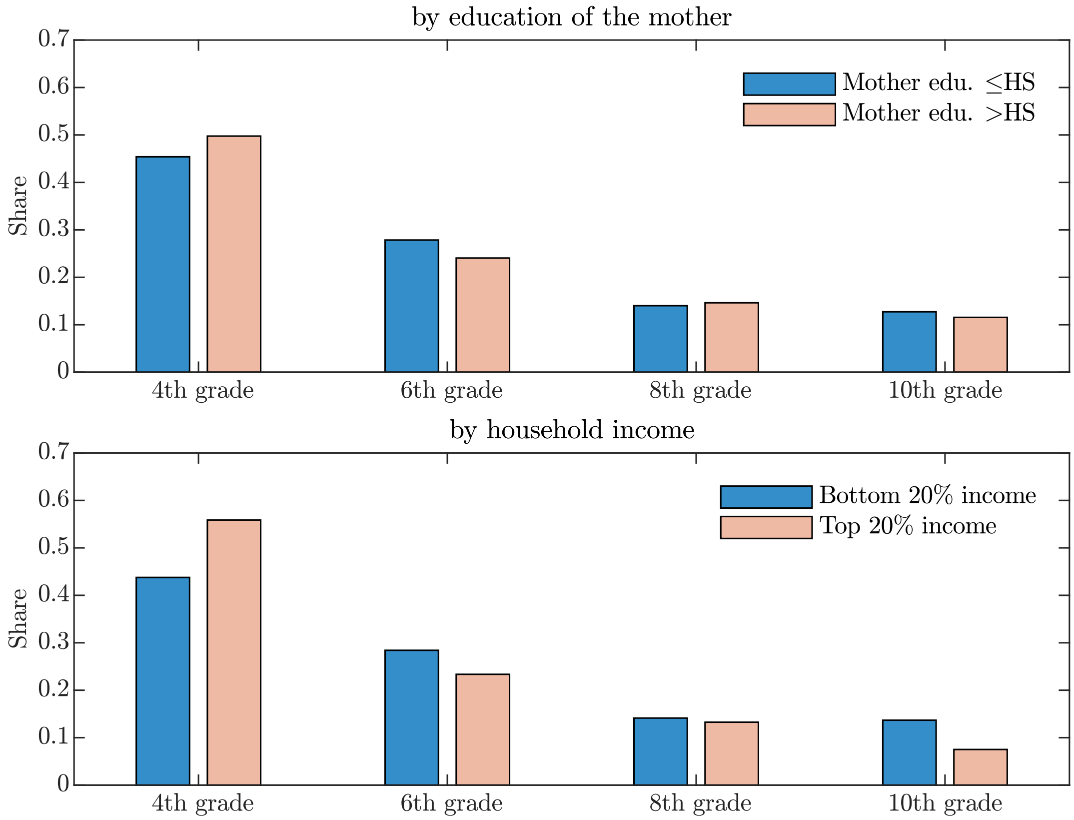
Note: The values on the graphs are calculated using the estimates of Table S.1.8. The top and bottom panels report parental and students responses, respectively. The vertical lines are school network-clustered bootstrapped 95% confidence intervals. Symbols on top (x) indicate that the difference in responses between female and male students is statistically significant at 1%. The values of graphs in the top panel represent weekly hours, while the values at the bottom of each plot indicate the response as a percent relative to the average time investment at each specific grade. Back to Section 1.7.3.

Figure S.2.12:
Model fit of targeted moments



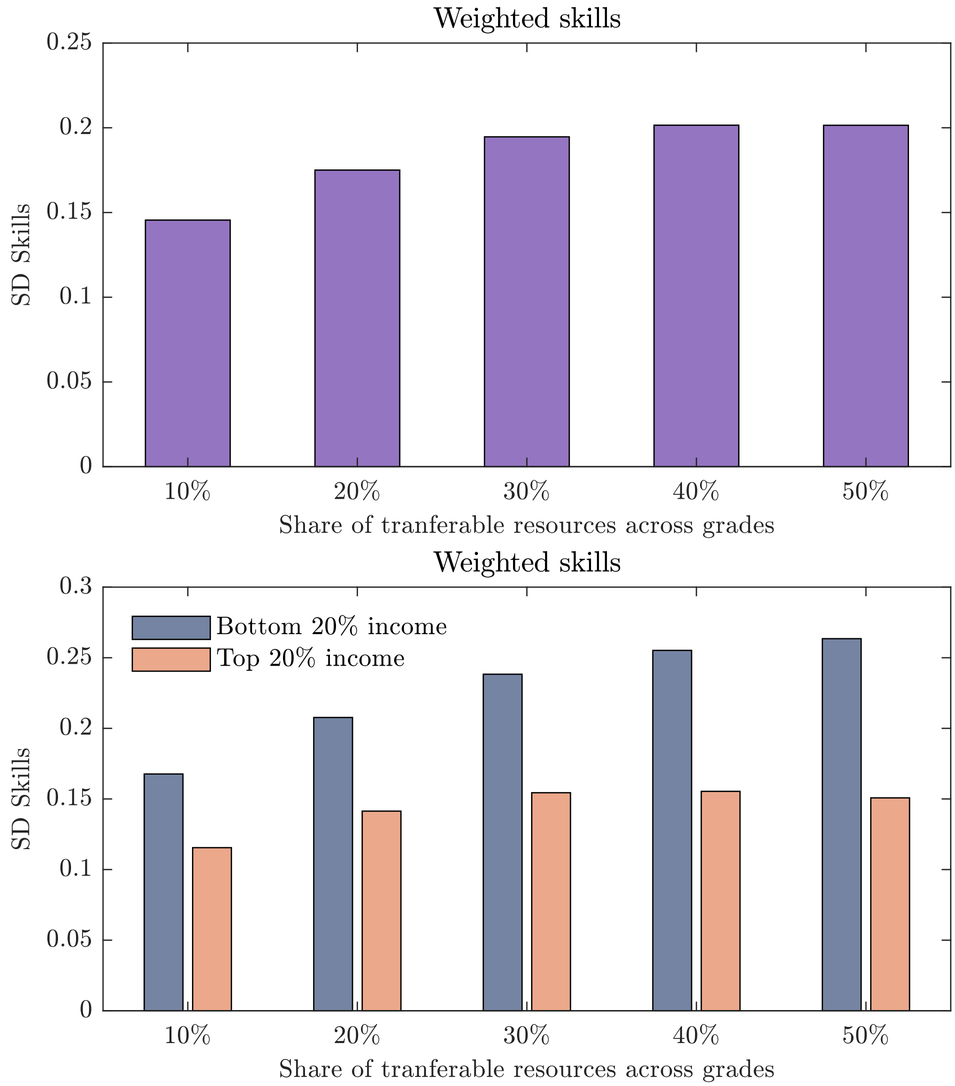
Note: Each dot in the graphs represents a moment used in the auxiliary model of the indirect inference estimator. The horizontal axis shows the value of the moment estimated in the data and the vertical axis shows the value from the data simulated with the child development model. Back to Section 2.2.4.

Figure S.2.13:
Optimal resource allocation across grades
Average shares of transferable resources by students' characteristics



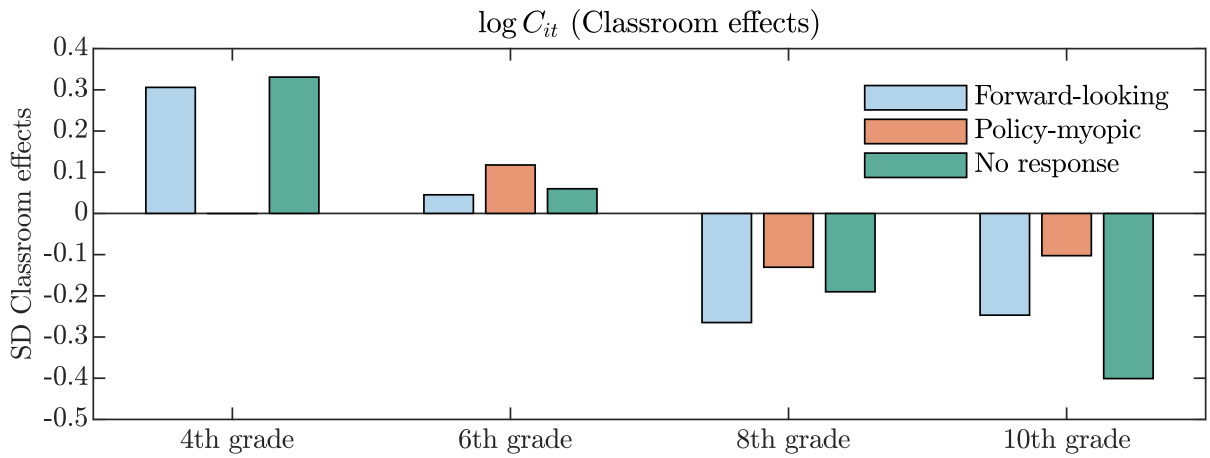
Note: This figure shows the average optimal shares of transferable resources at each school grade by mother education level and household income. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the child development model, I simulate the household choices and skills under each possible resource allocation across grades for each student. The optimal allocation is given by the shares of total transferable resources at each grade that maximize the weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17, 0.24 and 0.49 for grades 4, 6, 8 and 10, respectively. In this exercise I allow 30 percent of the resources of the baseline allocation of each grade to be transferable across grades. See section 2.3.2 for additional details.

Figure S.2.14:
Difference in average weighted skills index between optimal and baseline allocation by share of transferable resources



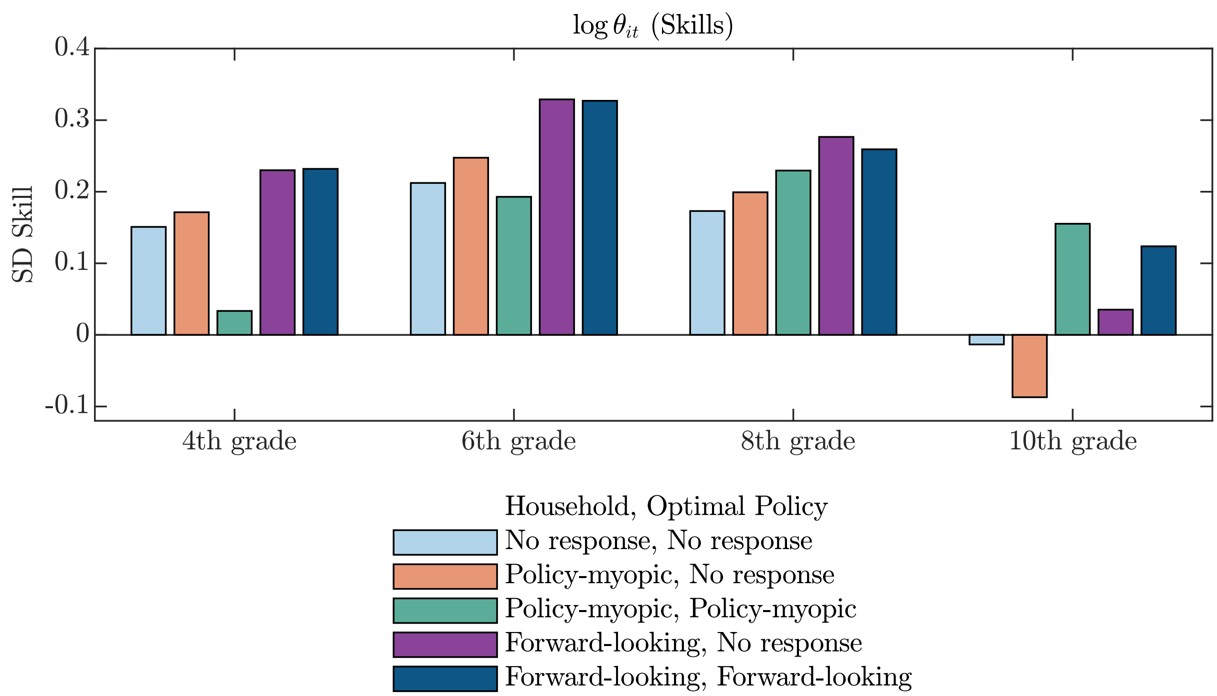
Note: The top panel presents the difference between the optimal and baseline allocation of the average weighted skills by fraction of transferable resources of each grade. The bottom panel shows the equivalent results for the top and bottom 20 percent of the household income distribution. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the child development model, I simulate the household choices and skills under each possible resource allocation across grades for each student. The optimal allocation is given by the shares of total transferable resources at each grade that maximize the weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17, 0.24 and 0.49 for grades 4, 6, 8 and 10, respectively. In this exercise I allow 30 percent of the resources of the baseline allocation of each grade to be transferable across grades. See section 2.3.2 for additional details.

Figure S.2.15:
Average differences in classroom effects between optimal and baseline allocation by household behavioral type



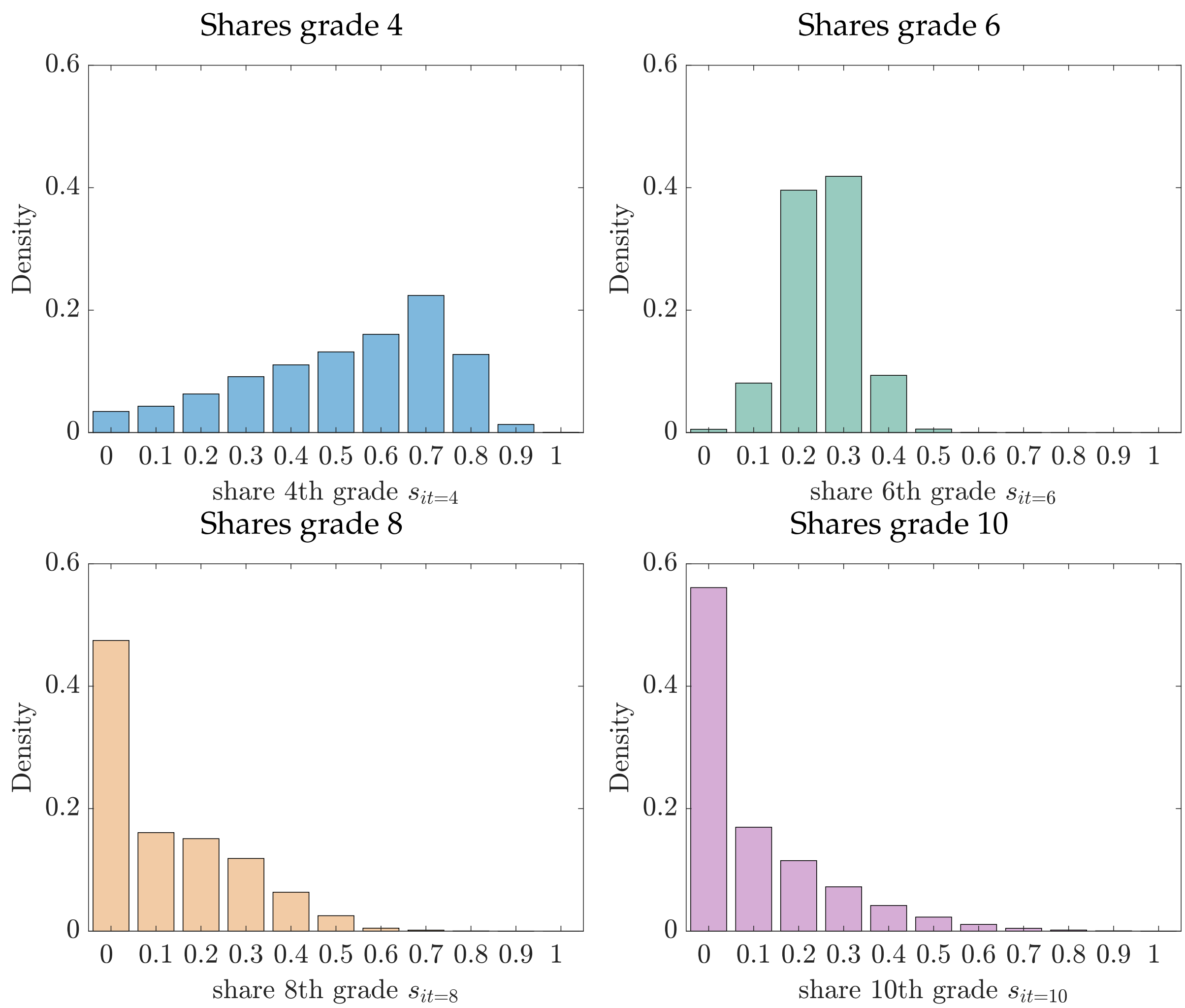
Note: The figure presents the difference of the average classroom effects between the optimal and baseline allocation by household type. No response: households do not respond to different allocations; Policy myopic: households respond as if resources in subsequent grades are the given by baseline allocation; Forward-looking: households are forward-looking and make their decisions understanding the dynamic implications of different allocations. For non-responsive households I simulate skills dynamics using the skill technology (Table S.1.5). For policy myopic households I simulate their choices and skills with the time investment functions (Table S.1.8) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 2.2.1). For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the child development model, I simulate the household choices and skills under each possible resource allocation across grades for each student. The optimal allocation is given by the shares of total transferable resources at each grade that maximize weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17,.0.24 and 0.49 for grades 4, 6,8 and 10, respectively. In this exercise I allow 30 percent of the resources of the baseline allocation of each grade to be transferable across grades. See section 2.3.2 for additional details.

Figure S.2.16:
Average differences in skills between optimal and baseline allocation
by household behavioral type



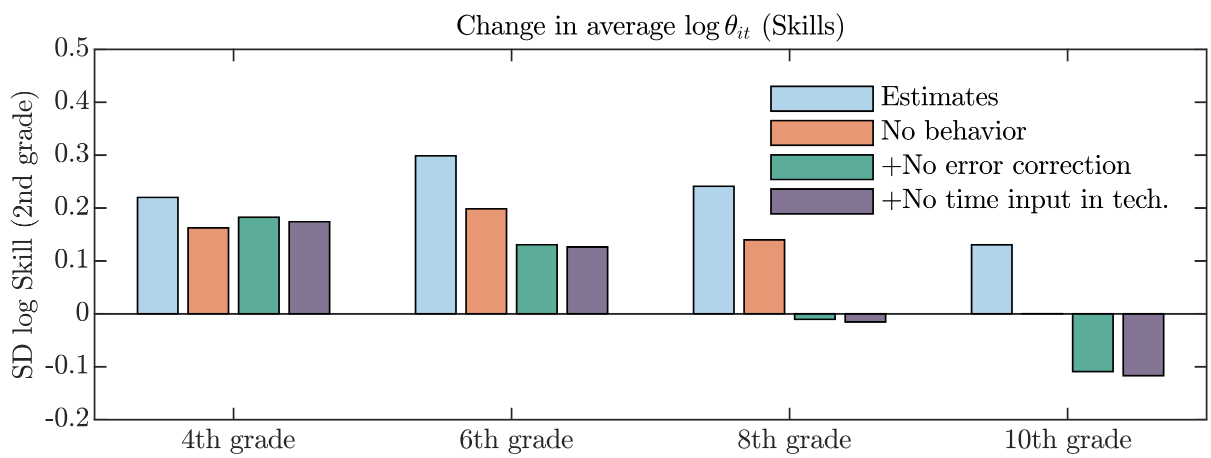
Note: The figure presents the difference of the average skills between the optimal and baseline allocation. The results are by household type (first word in bar's label) and by implementing the optimal allocation for a particular household type (second word in bar's label). Household types: No response: households do not respond to different allocations; Policy myopic: households respond as if resources in subsequent grades are the given by baseline allocation; Forward-looking: households are forward-looking and make their decisions understanding the dynamic implications of different allocations. For non-responsive households I simulate skills dynamics using the skill technology (Table S.1.5). For policy myopic households I simulate their choices and skills with the time investment functions (Table S.1.8) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 2.2.1). For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the child development model, I simulate the household choices and skills under each possible resource allocation across grades for each student. The optimal allocation is given by the shares of total transferable resources at each grade that maximize weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17,.0.24 and 0.49 for grades 4, 6,8 and 10, respectively. In this exercise I allow 30 percent of the resources of the baseline allocation of each grade to be transferable across grades. See section 2.3.2 for additional details.

Figure S.2.17:
Policy counterfactual: Distribution of optimal shares of transferable resources



Note: This figure shows results of policy that reallocates optimally transferable resources across grades 4 to 10. Each plot shows the distribution of optimal shares of total transferable resources assigned at each grade. Note that at each plot the bars sum up to one. See section 2.3.2 for additional details.

Figure S.2.18:
Optimal resource reallocation across grades
Estimates with and without error correction



Note: The figure shows the difference in average skills at each grade between the optimal and baseline allocation of transferable resources. The bars labeled "Estimates" simulate the skills under the full model estimates. The bars labeled "No behavior" show the estimated impact under the assumption that households do not respond—i.e., I simulating the skills under the policy only with the skill technology (Table S.1.5). Lastly, the bars labeled "+No error correction" and "+No time inputs in technology" simulate the skills with only the skill technology without correcting for measurement error and additionally not including time inputs, respectively. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (2.2.4). Then, with the child development model, I simulate the household choices and skills under each possible resource allocation across grades for each student. The optimal allocation is given by the shares of total transferable resources at each grade that maximize weighted skills index of each student. The weighted skills are a weighted average of the skills across grades 4 to 10 and the weights correspond to the coefficients of a regression of college attendance on measures of skills. The weights (normalized to sum up to one) are 0.10, 0.17, 0.24 and 0.49 for grades 4, 6, 8 and 10, respectively. In this exercise I allow 30 percent of the resources of the baseline allocation of each grade to be transferable across grades. See section 2.3.2 for additional details.