Abstract

This paper studies the relationship between school and home investments in the cognitive development of children and the behavior of the actors involved in the process. I employ large-scale administrative and survey data from Chile to estimate how parents and children in primary and secondary school adjust their time investment in response to classroom and teacher quality. Since classroom inputs are not directly observed, I first estimate the production function of cognitive skills that provides measures of classroom and teacher quality. I then estimate the time investment responses to these quality measures. I find that parents of younger children compensate for low quality whereas parents of older children reinforce quality. Students, on the other hand, increase time self-investment in response to higher quality, but the responses are larger for older children. Motivated by the heterogeneity in responses by school grade, I estimate a child development model and characterize the optimal allocation of school resources across grades. I find that it is optimal to allocate relatively more resources to lower grades. Moreover, ignoring the behavioral response of households implies an optimal allocation with substantially lower improvements in cognitive development.
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 the behavioral response to classroom inputs have implications on school resource allocation, teacher selection, and pay-for-performance policies.

In this paper, 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 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.

---

\(^1\)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 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 Strom and Falch (2020) for surveys on teacher value-added and estimation methods.
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.\footnote{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.} 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.\footnote{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).} 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.

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 secure exclusionary 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 procedure provides the estimate of 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 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.

A dynamic skill formation technology and the differential response of household to classroom quality by grade 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 response 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 estimates vary with changes in the expected classroom environments of subsequent grades. To evaluate the implications of different allocation 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 given by 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 auxil-
iary 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 paper 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 strand 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. These papers study responses to specific classroom components or proxies of school quality.\(^5\)

\(^5\)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
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 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. 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. 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 in the black box of classroom and teacher value-added and quantify the behavioral component in these measures of effectiveness. This information can be use in the design school resource allocation and teacher-related policies.

(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.

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.

An exception is Greaves et al. (2019) that finds that children increase effort in light of positive information about school quality.
Lastly, the paper contributes by characterizing the optimal allocation of school resource 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 paper is as follows. Section 2 develops a child development model and specifies the household problem, the skill formation technology and time investments functions. Section 3 presents the data and descriptive statistics. Section 4 describes the measurement systems of skills and time investment. Sections 5, 6, and 7 present identification, estimation methodology and estimates, respectively. Each section addresses measurement systems, skill technology, and time investments functions. In Section 8, I specify the parameterization of preferences and the expectations process, the estimation methodology of the child development model and estimates. Section 9 consists of the policy counterfactual analysis of the optimal resource allocation across grades. Finally, Section 10 concludes the paper.

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.\footnote{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). Alborno et al. (2018) develop a theoretical model that analyzes the interaction between student, parents and teachers.}
2.1 Household problem

The model consists of a unitary household that consists of parents and a single child. Throughout the paper the index $i$ is used interchangeably between child and household and the index $t$ between 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.\footnote{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.} Let $\Omega_{it} = \{\theta_{it}, C_{it}, x_{it}, z_{it}\}$ be the state space, where $\theta_{it}$ is the child’s cognitive skill, $C_{it}$ is classroom inputs, and the vectors $x_{it}$ and $z_{it}$ are exogenous characteristics. The vectors $x_{it}$ and $z_{it}$ might share elements and they influence preferences and the skill formation technology, respectively. The instantaneous utility function is denoted by $u_{it}$ and its arguments are the child’s skill stock at the beginning of the period $\theta_{it}$, time investment $h_{it}$ (vector size $D$), and the vector $x_{it}$. 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 $\beta$ is the discount factor. To simplify notation, I omit the $i$ index in this section. The problem of the household is:

\begin{equation}
V_t(\Omega_t) \equiv \max_{h_t} \left\{ u_t(\theta_t, h_t, x_t) + \beta E[V_{t+1}(\Omega_{t+1})] \right\}
\end{equation}

subject to

\begin{align*}
\text{Time constraint} & \quad h_t \in [0, H]^D; \\
\text{Production function} & \quad \theta_{t+1} = F_t(\theta_t, h_t, C_t, z_t, v_t),
\end{align*}

(2.1)

where the control variable $h_t$ is the time the household invest in the child. I assume that, conditional on $\Omega_t$, the household does not choose a classroom—i.e., classroom inputs are exogenous.\footnote{A more general model can incorporate school choice and this decision affects 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; or at least in terms of the expected quality of the school inputs. Thus, classroom inputs exogeneity, conditional on $\Omega_t$, is arguably a weak assumption.} 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 $z_t$, and an unobserved (to the household and econometrician) shock $v_t$. 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 \( \nu_t \) and future classroom inputs.

Note that time investment \( h_t \) can be multidimensional. For example, time investment provided by different household members. 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 children while the latter is the time children spend (by their own) studying, doing homework, or 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 in order to simplify notation and facilitate tractability—i.e., \( D = 1 \). 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:

\[
\frac{\partial u_t(h_t, \theta_t, x_t)}{\partial h_t} + \beta \frac{\partial E[V_{t+1}(\Omega_{t+1})]}{\partial h_t} = 0. \tag{2.2}
\]

The first term of equation (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 (2.2)—i.e., the policy function—can be expressed as:

\[
h_t = h_t^f(\theta_t, C_t, x_t, z_t). \tag{2.3}
\]

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 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 (assuming no uncertainty) for the sake of exposition, the partial derivative of the marginal benefit with respect to classroom inputs is:

\[
\frac{\partial^2 V_{t+1}(\Omega_{t+1})}{\partial \theta_t^2} \frac{\partial F_t}{\partial C_t} \frac{\partial F_t}{\partial h_t} + \frac{\partial V_{t+1}(\Omega_{t+1})}{\partial \theta_{t+1}} \frac{\partial^2 F_t}{\partial h_t \partial C_t}. \tag{2.4}
\]
The interpretation of equation (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 $\partial^2 F_i / \partial h_i \partial \theta_i < 0$, these inputs are substitutes in production, higher levels of classroom quality are associated with lower productivity of time investment. Instead, $\partial^2 F_i / \partial h_i \partial \theta_i > 0$ implies complementarity in production, and classroom quality makes the households investment more productive.\(^\text{12}\) The overall sign depends on preferences and on the skill technology, and ultimately, this is an empirical question.\(^\text{13}\)

### 2.2 Skill formation technology

This section describes the dynamics of the skill accumulation process. The cognitive skills of a child $\theta_{it}$ follows 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:

$$\theta_{it+1} = F_i(\theta_{it}, h_{it}, C_{it}, z_{it}, v_{it}), \quad (2.5)$$

where $F_i(\cdot)$ is a grade-specific function that depends on current skill, $\theta_{it}$, the time the household invest in its child $h_{it}$, the classroom inputs $C_{it}$, household characteristics $z_{it}$, and the structural shock $v_{it}$.

#### 2.2.1 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.\(^\text{14}\) This parametrization is flexible in terms of the relationship between inputs in the production

\(^{12}\)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.

\(^{14}\)In the data, I observe a non-trivial fraction of households choosing to invest zero time in the skill formation process.
of skills—i.e., it is possible a negative cross-derivative (see Agostinelli and Wiswall, 2016). It allows complementarity or substitutability between time and classroom inputs. The parametric functional form is the following:

$$\log \theta_{it+1} = \log F_t(\theta_{it}, h_{it}, C_{it}, z_{it}, \nu_{it}) = \log A_t + \gamma_1 \log \theta_{it} + \gamma_2 h_{it} + \gamma_3 \log C_{it} + \gamma_4 h_{it} \times \log C_{it} + z_{it}' \gamma_5 + \nu_{it},$$

where $A_t \exp(z_{it}' \gamma_5)$ is the total factor productivity. The set $\{\gamma_j\}_{j=1}^5$ defines the elasticity or semi-elasticity of next grade’s skills and inputs and $\nu_{it}$ is a mean zero shock.

The specification is more general in the empirical implementation. Besides the terms in equation (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 investment. However, equation (2.6) reduces notation burden and the identification and estimation analysis under this simplification is without loss of generality.

### 2.2.2 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 Hanushek, 2020). In particular, there is an extensive line of research studying the contribution of teachers to 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 by 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 $r_{it}$ and $\xi_{it}$ are the observed and unobserved (to the econometrician) classroom inputs, respectively. The technology of the

---

15Assuming a Cobb-Douglas or Constant Return to Scale production function with standard parameters values imply 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.

16In 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.
classroom inputs is:

\[ \log C_{it} = \psi_{1t} \log T_{it} + r'_{it} \psi_{2t} + \xi_{it}, \]  

(2.7)

where \( \psi_{1t} \) and \( \psi_{2t} \) represent each of these components’ contribution to the classroom effects. Under this specification, the skill formation technology has the following structure:

\[ \log \theta_{it+1} = \log A_t + \gamma_{1t} + \gamma_{2t} h_{it} + \gamma_{3t} \log T_{it} + r'_t \gamma_{4t} \]

\[ + \gamma_{6t} h_{it} \times \log T_{it} + h_{it} \times r'_t \gamma_{7t} + z'_t \gamma_{8t} + \nu_{it}. \]  

(2.8)

It should be noted that the error term \( \nu_{it} \) includes unobserved classroom level inputs \( \xi_{it} \)—in contrast with equation (2.6). Thus, identification of these parameters requires stronger assumptions than the one in equation (2.6).

### 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 \( h_{it} = h^f_{it}(\theta_{it}, C_{it}, x_{it}, z_{it}) \). It is a function of current skill, classroom inputs and 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 an specification that represents an approximation of the non-parametric policy function:

\[ h_{it} = \delta_{0,t} + \delta_{C,t} \log C_{it} + \delta_{\theta,t} \log \theta_{it} + \Gamma'_{\theta,t} \delta_{\Gamma,t} + \pi_i + \eta_{it}, \]  

(2.9)

where \{\( \delta_0, \delta_{C,t}, \delta_{\theta,t}, \delta_{\Gamma,t} \)\} are parameters, \( \Gamma_{it} \subset \{x_{it}, z_{it}\} \) is a vector of time-varying demographic characteristics, \( \pi_i \) is a household idiosyncratic component and \( \eta_{it} \) is a disturbance term. Under the specification that includes within classroom components, the time investment function is:

\[ h_{it} = \delta_{0,t} + \delta_{T,t} \log T_{it} + \delta_{\theta,t} \log \theta_{it} + r'_t \delta_{r,t} + \Gamma'_{\theta,t} \delta_{\Gamma,t} + \pi_i + \eta_{it}. \]  

(2.10)

\(^{17}\)An idiosyncratic component of the utility function implies that the function \( h^f_{it}(\cdot) \) is indexed by \( i \).
In equation (2.10) I allow for household responses to teachers (log $T_{it}$) to be different than to responses to other observable classroom inputs ($r_{it}$). Similar to the case of the skill technology, the required identification assumptions over $\eta_{it}$ are stronger, since there are additional parametric assumptions on the technology of classroom inputs and potential unobserved classroom inputs.\textsuperscript{18} If equations (2.9) and (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 the movements across the distribution of classroom inputs—e.g., moving a student from the 25th to 75th percentile in the classroom quality distribution.

Estimating equation (2.9) and (2.10) provide estimates of the responses of household without the need of specifying and estimating the full household model in equation (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 preference and beliefs and estimate the model which allows them to evaluate counterfactual policies that require household to update expectations. In this paper, I follow both approaches. I estimate the reduced form (in the literal sense) given by equation (2.9) and (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 (2.1), which allows me to evaluate a broader policy-space.

3 Data and descriptive evidence

3.1 Data

In this paper, 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 SIE 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. It has tests 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). Table 1 indicates for which grades and years where there is information available for the time period relevant for the analysis. 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. This property implies that the test score scale is invariant to affine transformations, but not invariant to monotonous 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

---

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. In 2008 there was a large voucher program that change significantly the school market (Neilson, 2013). Thus, using information post 2011 helps avoiding 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.
household responses does not depend on the scale of test scores, but the scale’s properties plays a key role in comparing the effects of inputs on cognitive 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 students performance and home investments measures, making the SIMCE data well suited to study the relationship between classroom inputs and time investment of households.23

I use two additional sources of survey data. One is the Time Use Survey of Chile, which is a national 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 last week 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 the test scores of 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 of children at each age while both surveys are representative of the same population.24 Appendix A describes the data in additional detail.

23Bharadwaj 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.

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

The estimates involving teachers require additional sample restrictions. I drop students whose math teachers I do not observe in at least two years in their careers. Table 3 shows the characteristics of math teachers in this sample. They are on average around 40 years old at every school grade. 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

---

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 turned 7 years old.

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.
more teaching experience and higher tenure at the school than teachers in upper grades.

Table 4 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 2 shows the distribution for both time investments by school grade.

### 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. 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, 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. 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.

Figure 1 plots the coefficients of the regressions. Each dot represents the coefficient using

---

29 Table 5 reports a subset of questions and their structure translated to English.

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:

$$A_{ijt} = X_{ijt}'\beta + V_{jt} + \epsilon_{ijt}$$

where $X_{ijt}$ is a vector of the student’s characteristics including a second order polynomial on 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 $\epsilon_{ijt}$ is an error term. The estimates of $V_{jt}$ are the measures of classroom value-added.

31 I regress each time investment measures 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.
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
development of the time investment question. These numbers are not comparable 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
top 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 bottom 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.

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 relevance. 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.

4 Measurement

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.

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:

$$M_{itm} = \mu_{itm} + \lambda_{itm} \log \theta_{it} + \epsilon_{itm}$$  (4.1)
where $\theta_{it}$ is the skill level of child $i$ at school grade $t$, $\mu_{tm}$ and $\lambda_{tm}$ are the parameters of the measurement system and $\varepsilon_{itm}$ is a zero mean error term. Assuming that the measures are linearly related to the natural logarithm of skills constrain the skill values to be positive numbers, which is required given the functional form assumed for the skill formation technology in equation (2.6). The parametric assumption of the technology and its implications for the measurement system is 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.

### 4.2 Time investment measurement

Existing work has used linear and non-linear measurement systems for latent investments. 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 build 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 about 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.

In this paper, 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 to incur in 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. A nice feature of the non-linear approach is that it

---


33The 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.
estimates time investment of each student measured in hours, as opposed to standard deviations, leading to a more clear interpretation of 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 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:

\[
Z_{its} = k \quad \text{if and only if} \quad \alpha_{stk} \leq \beta_{st} h_{it} + \epsilon_{its} < \alpha_{stk+1}
\]

(4.2)

for \( k = 1, 2, \ldots, K_s \)

where \( \alpha_{sl1} = -\infty \) and \( \alpha_{sk_{s+1}} = \infty \). I assume that the error terms \( \epsilon_{its} \) 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.

5 Identification

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 5 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 \( \beta_{st} \) 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.
5.1 Measurement system

I build a measurement system for cognitive skills of children and 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 by parents and children in the administrative data.

5.1.1 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. 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, 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—

---

37I 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 measures children’s cognitive development and track its evolution as they grow up. That is, 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 test are design differently at every grades. However, I follow this nomenclature to facilitate the comparison between the two kinds of measures.
i.e., equation (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 $M_{itm}^A$ be an 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:

$$M_{itm}^A = \mu_m^A + \lambda_m^A \log \theta_{it} + \epsilon_{itm},$$

where $\theta_{it}$ is the skill of the child and $\epsilon_{itm}$ is a mean zero error term. Note that the parameters $\mu_m^A$ and $\lambda_m^A$ 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:**
- $\epsilon_{itm} \perp \epsilon_{itm}'$ for all $t$ and all $m \neq m'$;
- $\epsilon_{itm} \perp \log \theta_{it}$ 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 $\theta_{it}$. That is:

**Normalization 1:**
- $E(\log \theta_{i0}) = 0$;
- $\text{Var}(\log \theta_{i0}) = 1$.

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:

\[ \mu^A_m = E(M^A_{i0m}); \quad \lambda^A_m = \sqrt{\frac{\text{Cov}(M^A_{i0m'}, M^A_{i0m}) \text{Cov}(M^A_{i0m}, M^A_{i0m'})}{\text{Cov}(M^A_{i0m'}, M^A_{i0m'})}}. \]  

(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 \( t \):

\[
E(\log \theta_{it}) = \frac{E(M^A_{itm}) - \mu^A_m}{\lambda^A_m}; \quad \text{Var}(\log \theta_{it}) = \frac{\text{Cov}(M^A_{itm}, M^A_{itm'})}{\lambda^A_m \lambda^A_{m'}}. 
\]

(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:

\[
M_{itm} = \mu_{mt} + \lambda_{mt} \log \theta_{it} + \varepsilon_{itm},
\]

(5.4)

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

\[
\lambda_{mt} = \sqrt{\frac{1}{\text{Var}(\log \theta_{it})} \times \frac{\text{Cov}(M_{itm}, M_{itm'}) \text{Cov}(M_{itm}, M_{itm''})}{\text{Cov}(M_{itm'}, M_{itm''})}}, \quad \mu_{mt} = E(M_{itm}) - \lambda_{mt} E(\log \theta_{it}).
\]

(5.5)

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

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

(5.6)

It is important to note that the operation in equation (5.6) is an affine transformation of the test scores. Thus, the original scale of the test scores is preserved.\(^{38}\)

\(^{38}\)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.
5.1.2 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 $g_t(\cdot)$. 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:
- $\epsilon_{it} \perp h_{it}$ for all $s$ and all $t$;
- $\epsilon_{it} \perp \epsilon_{it'}$ 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 $s$: $\alpha_s$ and $\beta_s$. The expected value of observing outcome 1 for measure $s$ is:

$$p_s \equiv \Pr(Z_{is} = 1) = \int \Pr(Z_{is} = 1 \mid h_{i}; \alpha_s, \beta_s) g(h_{i}) dh_{i}, \quad (5.7)$$

Given $g(\cdot)$ and $p_s$, for any $\beta_s$ the function $\Pr(Z_{is} = 1)$ is strictly decreasing in $\alpha_s$, so the inverse with respect to $\alpha_s$ 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:

$$p_{s,s'} = \Pr(Z_{is} = 1, Z_{is'} = 1)$$

$$= \int \Pr(Z_{is} = 1 \mid h_{i}; \alpha(\beta_s, p_s), \beta_s) \times \Pr(Z_{is'} = 1 \mid h_{i}; \alpha(\beta_{s'}, p_{s'}), \beta_{s'}) g(h_{i}) dh_{i} \quad (5.8)$$

$$= q(\beta_s, p_s, \beta_{s'}, p_{s'}).$$

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

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.
\[ S \geq 3 \] — we have for \( s = 1, 2, 3 \):

\[
\begin{align*}
p_{1,2} &= q(\beta_1, p_1, \beta_2, p_2) \\
p_{1,3} &= q(\beta_1, p_1, \beta_3, p_3) \\
p_{2,3} &= q(\beta_2, p_2, \beta_3, p_3). \tag{5.9}
\end{align*}
\]

It follows that:

\[
p_{2,3} = q(\overline{q}(\beta_1, p_1, p_{1,2}, p_2), p_2, \overline{q}(\beta_1, p_1, p_{1,3}, p_3), p_3) \equiv r(\beta_1, p_1, p_2, p_3, p_{1,2}, p_{1,3}). \tag{5.10}
\]

Lastly, since \( r(\cdot) \) is strictly decreasing in \( \beta_1 \) (Appendix B presents the proof), it follows that the inverse of \( r(\cdot) \) on \( \beta_1 \) exists and so \( \beta_1 \) is identified. Similarly, \( \beta_s \) and \( \alpha_s \) are identified for all \( s \). Note that without knowledge of \( g_l(\cdot) \) 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 to identify the latent value of each individual 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) develop a different approach and he focused 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 largely used and extended (Mislevy, 1993; Junker et al., 2012; Schofield, 2014; Schofield, 2014).

\(^{40}\)In the same vein, Williams (2019) uses a semiparametric approach to identify the relationship of two variables while “controlling” for the latent variable.
et al., 2015). Mislevy proposed using plausible values but from the posterior distribution conditional on the ordered categorical questions and the variables which their relationship to the latent variable is of interest. 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 identify 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 to identify the latent value of each student as the mode of the individual likelihood times a weighting function as:

$$\tilde{h}_{it} \equiv \arg\max_{h_{it}} w(h_{it}) \times f(z_{it} \mid h_{it}; \alpha, \beta)$$

(5.11)

where $w(h_{it})$ is the weighting function. By definition, the derivative of the function of the right side of equation (5.11) is equal to zero at the mode:

$$\frac{\partial \log f(z_{it} \mid h_{it}; \alpha, \beta)}{\partial h_{it}} + \frac{\partial \log w(h_{it})}{\partial h_{it}} = 0.$$  

(5.12)

The first term of the equation (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 $B(h_{it})$, defined as difference between the mode of the individual likelihood and the time investment, times the Fisher information function.

---

4) Plausible values are defined as random draws from the latent variable’s posterior distribution. The plausible values (Rubin, 1987) difference with 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.
\[ I(h_{it}):^{42,43} \]
\[
\frac{\partial \log f(z_{it} | \tilde{h}_{it})}{\partial h_{it}} + B(\tilde{h}_{it})I(\tilde{h}_{it}) = 0. \tag{5.13}
\]

Then, Warm proofs 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 \((\alpha, \beta)\) are not observed, the identification of \( h_{it} \) is up to some error. I define this error as \( \zeta_{itl} \) and the index \( l \) indicates the set of variables \( Z_{it} \) included in the identification of \( h_{it} \).

### 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 (5.6) and (5.13) in equation (2.6) and rearrange:

\[
\tilde{M}_{it+1m} = \log A_t + \gamma_{it} \tilde{M}_{itm} + \gamma_{2t} \tilde{h}_{itl} + \gamma_{3t} \log C_{it} + \gamma_{4t} \tilde{h}_{itl} \times \log C_{it} + \tilde{Z}'_{it} \gamma_{5t} + \tilde{v}_{it}, \tag{5.14}
\]

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

\[
\tilde{v}_{it} \equiv v_{it} - \frac{\gamma_{1t} \tilde{v}_{itm}}{\lambda_{mt}} - \gamma_{2t} \tilde{\zeta}_{itl} = \gamma_{4t} \tilde{v}_{itl} \times \log C_{it} + \frac{\tilde{v}_{it+1m}}{\lambda_{mt+1}}. \tag{5.15}
\]

There are two threats to identification of the parameters in equation (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 \( v_{it} \). 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 preference. Moreover, since classroom effects do not have a natural scale or location, I perform the following normalization:

**Normalization 2:**

- \( E[\log C_{it}] = 0 \) and;

\[ ^{42} \text{Warm (1989) considered dichotomous ability measures and Samejima (1993) provided an extension of the second term in Warm’s equation (5.12) for polytomous measures, which is the case of the measures used in the current setting.} \]

\[ ^{43} \text{Even though } w(h_{it}) \text{ does not have closed form, both } B(h_{it}) \text{ and } I(h_{it}) \text{ do have closed form; I provide the analytic functions in Appendix B.} \]
• $\text{Var}[\log C_{it}] = 1$ for every $t$.

The identification of the parameters under the presence of measurement error requires exclusionary 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 $d_{it} \in \{\log C_{it}, \log \theta_{it}, h_{it}, z_{it}\}$ and $\omega_{itj} \in \{\varepsilon_{itm}, \zeta_{its}\}$ where $j = m, s$.
- $\omega_{itj} \perp \omega_{itj'}$ for all $j$ and $t \neq t'$;
- $\omega_{itj} \perp \omega_{itj'}$ for every $t \neq t'$ and all $j$ and $j'$;
- $\omega_{itj} \perp \omega_{itj'}$ for all $\omega \neq \omega'$ and all $t$ and $t'$;
- $\omega_{itj} \perp d_{it}$ for all $j$ and all $t$;
- $\omega_{itj} \perp v_{it'}$ for all $j$ and all $t$ and $t'$.

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

**Assumption 4:** mean-independence:

$$E[v_{it} \mid \log C_{it}, \log \theta_{it}, h_{it}, z_{it}] = 0. \quad (5.16)$$

Then, under Normalization 1 and 2 and Assumption 1 to 4 the parameters of equation (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 Assumption 1 to 4 and Normalization 1 and 2, the classroom effects are identified up to some error. Since the classroom effects 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:

$$\log \tilde{C}_{ite} = \log C_{it} + \chi_{ite}, \quad (5.17)$$

where $\chi_{ite}$ is the error associated with a particular sample of students, indexed by $c$. 

27
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 (5.6), (5.13), and (5.17) into equation (2.9) and rearrange:

\[
\tilde{h}_{itl} = \delta_{0,t} + \delta_{C,t} \log \tilde{C}_{itc} + \delta_{\theta,t} \tilde{M}_{itm} + \Gamma_{it} \delta_{\Gamma,t} + \pi_i + \tilde{\eta}_{itl},
\]

where \( \tilde{\eta}_{itl} = \eta_{itl} - \delta_{C,t} X_{itc} - \delta_{\theta,t} \frac{\epsilon_{itm}}{\lambda_{tm}} + \zeta_{itl}. \) (5.18)

As in the case of the skill technology, there are two threats to identification of the parameters in equation (5.18): measurement error and selection in unobservables. The former is addressed by exploiting exclusion conditions using additional skill and classroom effects measures.\(^{44}\) These exclusionary restrictions hold if additionally to Normalization 1 and 2 and Assumptions 1 to 4, the following assumption is met:

**Assumption 5:** For \( d_{it} \in \{ \log C_{itl}, \log \theta_{itl}, h_{itl}, z_{itl} \} \) and \( \omega_{ij} \in \{ \epsilon_{itm}, \zeta_{itl} \} \) where \( j = m, l \).

- \( X_{itc} \perp \chi_{itc'} \) for all \( j \) and \( t \neq t' \);
- \( X_{itc} \perp \omega_{itj} \) for every \( t \) and \( t' \) and all \( c \) and \( j \);
- \( X_{itc} \perp d_{it} \) for all \( c \) and all \( t \);
- \( X_{itc} \perp \eta_{itl'} \) for all \( c \) and all \( t \) and \( t' \).

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

**Assumption 6:** mean-independence:

\[
E[\eta_{itl} | \log C_{itl}, \log \theta_{itl}, \Gamma_{itl}, \pi_i, t] = 0.
\] (5.19)

Under Normalization 1 and 2 and Assumptions 1 to 5 the parameters in equation (2.9) are identified. The proof is the usual one in the framework of fixed effects models with 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.

\(^{44}\)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.
6 Estimation

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.

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.

6.1.1 Skill measurement

The skill measurement system is estimated by replacing the moments in equations (5.2), (5.3) and (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 (5.6).

6.1.2 Time investment measurement

The questions of time investments 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:

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

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

$$
\Pr(Z_{its} = k \mid h_{it}) = \Pr(Z_{its} \geq k \mid h_{it}) - \Pr(Z_{its} \geq k + 1 \mid h_{it}).
$$
where \( \Pr(Z_{its} \geq 0 \mid h_{it}) = 1 \) and \( \Pr(Z_{its} > K \mid h_{it}) = 0 \). 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:

\[
 f(z_{it} \mid h_{it}; \alpha, \beta) = \prod_s \Pr(Z_{its} = z_{its} \mid h_{it}), \tag{6.3}
\]

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

\[
 L_{it}(\alpha, \beta \mid z_{it}) = \int f(z_{it} \mid h_{it}; \alpha, \beta) g_t(h_{it}) dh_{it}, \tag{6.4}
\]

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.\(^{45}\) There are two assumptions 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.\(^{46}\) 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 \( t \):

\[
 \log L_t(\alpha, \beta \mid z_{it}) = \sum_{i=1}^{N_t} \log L_{iti}(\alpha, \beta \mid z_{it}), \tag{6.5}
\]

\(^{45}\)In item response theory, where the latent variable is student’s 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 in the Chilean Time Use Survey.

\(^{46}\)The administrative data on time investments is reported through questionnaires to parents and students, which is not mandatory.
and the maximum likelihood estimator is defined as:

\[
(\hat{\alpha}, \hat{\beta}) \equiv \arg\max_{(\alpha, \beta)} \log L_i(\alpha, \beta \mid z_{it}).
\] (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 used to estimate each household’s time inputs. The item response theory developed 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 bias estimators and 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 to spread the estimates’ distribution relative to the distribution of the true latent variable across individuals.

Existing work that use non-linear measurement models for time investments, such as Agostinelli (2018) and Wang (2020), rely 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.\(^{47}\) Instead, Wang (2020) uses a multivariate ordered response model similar to the one presented in sections 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

\(^{47}\)In item response theory this kind of estimators of individuals’ latent variables are usually called plausible values (Rubin, 1987).
in consequence put some weight on the prior distribution, so the posterior distribution is compressed towards the expected value of 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 individual’s parental time, 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 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 student heterogeneity’s 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 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 in equation (5.13) and solving for \( \tilde{h}_{itl} \) for each student. That is,

\[
\frac{\partial \log f(z_{itl} \mid \tilde{h}_{itl}; \hat{\alpha}, \hat{\beta})}{\partial \tilde{h}_{itl}} + B(\tilde{h}_{itl}; \hat{\alpha}, \hat{\beta})I(\tilde{h}_{itl}; \hat{\alpha}, \hat{\beta}) = 0. \tag{6.7}
\]

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 \( \zeta_{itl} \). I estimate several times the response model using disjoint sets of the categorical outcome questions \( Z_{its} \). Each of these estimates provide time investment measures that have different estimation error \( \zeta_{itl} \), where each disjoint set of questions is indexed by \( l \). It follows that \( \zeta_{itl} \perp h_{itl} \) for all \( t \) and \( l \) and \( \zeta_{itl} \perp \zeta_{itl'} \) for all \( l \neq l' \) and all \( t \).
6.2 Skill formation technology

In this section, I describe the methodology 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 Arcidiacono et al. (2012)’s algorithm 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 on inputs. The methodology estimates classroom effects as the systematic variation in 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 $\gamma_0 \equiv (A_0^0, \gamma_1^0, \ldots, \gamma_4^0, \gamma_5^0)$. Each iteration $n \in \{0, 1, \ldots\}$ of the algorithm consist of computing the following steps:

**Step 1:** Taken as given the current parameter guess $\gamma_n$, compute the classroom effect as the average within-classroom residual in skills at grade $t$:

$$
\log C_{it}^n = \frac{\sum_{i' \in c(i)} \left[ M_{it+1m} - \log A_i^n - \gamma_1^n M_{itm} - \gamma_2^n \tilde{h}_{it} - z_i' \gamma_3^n \right]}{\sum_{i' \in c(i)} \left[ \gamma_3^n + \gamma_4^n \tilde{h}_{it} \right]} \quad (6.8)
$$

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

**Step 2:** Taken 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:

$$
\tilde{M}_{it+1m} = \log A_i^{n+1} + \gamma_1^{n+1} \tilde{M}_{itm} + \gamma_2^{n+1} \tilde{h}_{it} + \gamma_3^{n+1} \log C_{it}^n + \gamma_3^{n+1} \tilde{h}_{it} \times \log C_{it} + z_i' \gamma_3^{n+1} + \tilde{v}_{it} \quad (6.9)
$$

I estimate these parameters with 2SLS estimator using additional measures of skills and time inputs in the first stage.\textsuperscript{48} 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 $\gamma_t^{n+1}$. Once

\textsuperscript{48}The OLS estimation of the **Step 2** equation produces inconsistent estimates of the remaining parameters due to measurement error bias.
convergence is achieved, it provides the classroom effects distribution. I run the estimation by (disjoint) random samples of 50 percent of the students in each classroom to estimate classroom effects with different estimation error to exploit exclusionary restrictions in the estimation of the time investment functions.\footnote{Note that there is an implicit assumption that \( v_{it} \) is not observed by parents and children when they are making their investment choices. This can be relaxed by allowing correlation between \( v_{it} \) and \( \eta_{it} \) and using a control function approach as in Attanasio et al. (2020b). Under this framework, the estimation of skill technology and time investment functions is done in a single step—i.e., including the estimation of the skill technology in Agostinelli et al. (2020)’s estimator. The algorithm is initialized by assuming a guess of the parameters and of the distribution of \( \eta_{it} \) 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 \( \eta_{it} \) that can be use in the subsequent iterations until convergence is achieved.}

The algorithm can be modified to estimate the effects of different components of the overall classroom effect, such as the teacher effect \( \log T_{it} \) and 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.

### 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 (5.13), equations (5.6) and (5.17) into equation (2.9) and rearrange:

\[
\tilde{h}_{it} = \delta_{0,t} + \delta_{C,t} \log \tilde{C}_{it} + \delta_{\tilde{M},t} \tilde{M}_{it} + \Gamma_{it}' \delta_{l,t} + \pi_i + \tilde{\eta}_{it}
\]  

There are three concerns that could bias the estimates of the parameters of equation (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 $\eta_{it}$ is not correlated with the observed inputs, the OLS estimates of $\delta_{C,t}$ and $\delta_{0,t}$ could be biased due to measurement error. I use the additional measures of each of these variables to exploit exclusionary restrictions to estimate the parameters of equation (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. 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 changing covariates, such as household income. Thus, I estimate equation (6.10) using a 2SLS estimator with student fixed effects.

Since coefficients are grade-specific, I rewrite equation (6.10) with a vector of variables in fourth grade and the variables in upper grades interacted with a school grade indicator. Define the vectors $X_{it} = (\log \tilde{C}_{it}, \tilde{M}_{itm}, \Gamma_{it}')'$ and $\delta_t = (\delta_{C,t}, \delta_{0,t}, \delta_{1,t}')'$. The vector of variables to include in the 2SLS estimator is:

$$b_{it} = (1, X_{it}, \{1(t = t'), 1(t = t') \times X_{it}\}_{t' > 4}')'$$

with $t' = 4, 6, 8, 10$ and $1(t = t')$ as an indicator function equal to 1 if $t = t'$ and zero.

---

50 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 the student was 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.

51 Controlling for household income could be problematic. The parents’ earnings might be affected by the change in parental time. However, the results are not quantitative or qualitative 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.
otherwise. Similarly, I define the vector of associated coefficients:

\[
\delta = (\delta_{0,4}, \delta_4, \{\tilde{\delta}_{0,t'}, \tilde{\delta}_{t'}\}_{t'>4}, \delta_4', \delta_{0,4}', \tilde{\delta}_{0,t'}, \tilde{\delta}_{t'})',
\]

(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:

\[
\tilde{a}_{it} = a_{it} - \bar{a}_i + \bar{a},
\]

(6.13)

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

7 Estimates

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 bootstrap. 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.

7.1 Measurement systems

Figure 3 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 (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 6 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. Table 7 shows the estimates and signal share for the measurement system estimated using the test scores of the administrative data. The signal share of these measures is similar to previous work using other databases.

7.2 Skill formation technology

Tables 8 and 9 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. Consider the parental time’s marginal effect of a particular student:

$$\frac{\partial \log \theta_{it+1}}{\partial h_{it}} = \gamma_{3i} + 2\gamma_{4i} h_{it} + \gamma_{9i} \log C_{it} + \gamma_{11i} e_{it} + \gamma_{12i} \log \theta_{it}. \quad (7.14)$$

The parameters’ sign in the equation (7.14) define how inputs relate in production with parental time. For example, if $\gamma_{9i} > 0$, the effect of parental time is higher for children with better classroom inputs. 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, eighth, and tenth grade there is

---

52 The definition of the signal share is $1 - \text{Var}(e_{it|m})/\text{Var}(M_{it|m})$. See equation (4.1) for more details.

53 The notation of this specification is as follows:

$$\log \theta_{it+1} = \log A_i + \gamma_{1i} \theta_{it} + \gamma_{2i} \log^2 \theta_{it} + \gamma_{3i} h_{it} + \gamma_{4i} h_{it}^2 + \gamma_{5i} e_{it} + \gamma_{6i} e_{it}^2 + \gamma_{7i} \log C_{it} + \gamma_{8i} \log \theta_{it} \times \log C_{it} + \gamma_{9i} h_{it} \times \log C_{it} + \gamma_{10i} e_{it} \times \log C_{it} + \gamma_{11i} h_{it} \times e_{it} + \gamma_{12i} h_{it} \times \log \theta_{it} + \gamma_{13i} e_{it} \times \log \theta_{it} + \epsilon_{it}' \gamma_{3i} + \nu_{it}.$$ 

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

54 Under the technology’s parametric assumption, this conclusion can only be reach if the sign is positive. If it is negative we could not derive the opposite conclusion. Formally:

$$\frac{\partial \log \theta_{it+1}}{\partial h_{it}} = \frac{\partial^2 \theta_{it+1}}{\partial h_{it} \partial C_{it}} \frac{\partial C_{it}}{\partial h_{it} \partial \theta_{it} \partial C_{it}} = \gamma_{9i}. \quad (7.14)$$

Thus, it could be that $\gamma_{9i} < 0$ and $\partial \theta_{it+1|m}/\partial h_{it} \partial C_{it} > 0$, 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 on the technology 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.
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 4 plots the sample average of equation (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 equation (2.6) have a log-log relationship with skills in the next grade. Hence, under the parametric assumption, the equivalent equation (7.14) for these inputs represent an elasticity. A 1 percent increase in classroom inputs increases between 0.56 and 0.44 percent skills. 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 comparing these effects across grades. There are two consideration 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. The effects seem large relative to the literature. The education production function literature finds that a 1 SD increase in the classroom distribution 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 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. Instead, second and tenth grades’ SD are similar (see Table 6). 2) the specifications in Tables 8 and 9 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 similar results than in the literature. I present these results in Table S1.

Table 4 shows the average level of time investments 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 relative more to rationalize the data.

---

55These effects seem large relative to the literature. The education production function literature finds that a 1 SD increase in the classroom distribution 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 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. Instead, second and tenth grades’ SD are similar (see Table 6). 2) the specifications in Tables 8 and 9 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 similar results than in the literature. I present these results in Table S1.

56Table 4 shows the average level of time investments 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 relative more to rationalize the data.
from 0.4 to 0.3 percentiles. The average effect of one weekly hour is larger for child self-
investment, it decreases monotonously from 1.5 to 0.6 percentiles from grades 4 to 10. 
Classroom effects in terms of percentiles have a U-shape form across grades, with values 
between 10.5 and 9 percentiles. Lastly, the average effect of current skills monotonously 
decreases from over 14 to 12 percentiles.\textsuperscript{57}

The technology estimates could be biased if there are unobservable inputs or character-
istics that systematically influence skills and inputs. On the one hand, if conditional on 
the classroom assignment there is selection bias in the estimates of the observable inputs’ 
effects. On the other hand, if the classroom effects are bias due to omitted variable bias.

The identification assumption relies in the rich data available on inputs and students’ charac-
teristics in 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 which 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 10 presents the results of this test. The first row shows the coefficients of regression 
of skills at each grade on household income—measured in 3,000 dollars, which represents 
around 30 percent of the standard deviation and average annual household income in 
the sample.\textsuperscript{58} An additional 3,000 dollars is associated with skills between 0.12 and 0.18 SD 
higher. Instead, the second row of the table shows the coefficient of regressing the estimated 
residuals of the skill technology on household income. In this case, additional 3,000 dollars 
in household income is associated with residuals between 0.0007 and 0.0037 SD (of skills)

\textsuperscript{57}The equivalent results for the specification with within classroom inputs are similar. These results are presented in Figures 6 and 7. 
\textsuperscript{58}Table 2 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.
higher. These estimates are statistically significant. However, given it is a large income change and small associated changes in residuals, the 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* prediction 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 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 extremely well the average skill of each household income decile even though this variable in not included in the specification.

Lastly, classroom effects are bias if the assignment of students to classroom is based on unobservable factors. To test this hypothesis, I regress the estimated classroom effects on household income. The results are in the first row of Table 11. An additional 3,000 dollars is associated with classroom effects between 0.04 and 0.09 SD larger. This suggests that students are assigned to classroom based on characteristics such as income. The second row of Table 11 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 and the relationship is not economically significant. This result suggests that most of the selection is through the school choice. 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 (6.10) aims to address this selection.

### 7.3 Time investment functions

This section describes the estimates of the time investment functions in equation (2.9). There are two different types of household time investments: parental and child time investment. Table 12 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.

---

59 In Chetty et al. (2014a)'s environment most of the selection is due to school choice as well, rather than across teachers.
specification of Table 8, while columns (3) and (4) show the responses to teacher effects associated with the estimated technology reported in Table 9. Note 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 9 and 10 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. Instead, 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., OLS estimator. The colored bars show the responses’ 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.

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. 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.

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 a hour per week at upper grades (Figure 10). These responses are large relative to the average child self-investment (between 17 and 7 percent, respectively). Meanwhile, the responses to teachers are smaller

---

60 The average time investment by school grade are presented in Table 4.
61 Figure S4 shows the responses to classroom and teacher reassignment between percentiles 10 and 90 across
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 be who teach most subjects and spend more time interacting with the students.62

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 decrease skills in 0.036 standard deviations (SD) of log skills in second grade. At tenth grade the parental response increases skills in 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.63

It is difficult to pin point the reasons of the change in the direction of 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 to 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 algebra.64 As children start learning subjects with more specialized knowledge—e.g., calculus—parents are less able to substitute teachers. However, they might incur in different activities, like advising, that could be more complementary to teaching quality in the formation of skills.65

62 Students interact, on average, with around 6 teachers in fourth grade and with almost 11 teachers in tenth grade (Table 2). Moreover, usually in lower grades there is a main teacher who is in charge on the “core” subjects, such as math, language, biology, and so on, while auxiliary teachers teach music, art and physical education.

63 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

64 This example is an extreme simplification of what teacher quality and responsibilities of teachers are.

65 Figure S1 presents results from the National Household Education Survey that attempt to provide suggestive evidence of this conjecture. It shows the share of parents that help with homework at 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 monotonous decline while the latter slightly increase as children grow up. This result suggest that parents might modify the composition of the activities considered as time investments as children grow up.
8 Model parameterization and estimation

The estimates in the previous section 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 constraint 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 2. The model allows evaluating policies that require households to update expectations of future classroom environments. I use the model to evaluate different allocation of resources across grades and characterize the allocation that maximizes the cognitive development of children.

8.1 Model parametrization

I depart from the standard utility function in terms of consumption. In the model, household value the cognitive skills of the child and it incurs in a disutility cost for each hour of parental and child time investment.\footnote{\textsuperscript{66}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 follow that specification.} I relabel parental and child time investment as $h_{it}$ and $e_{it}$, respectively. The utility functional form is:

$$u_{it}(\theta_{it}, h_{it}, e_{it}, x_{it}) = \frac{\theta_{it}^{1-\phi_{1i}} - 1}{1 - \phi_{1i}} - \phi_{2it} h_{it} - \phi_{3it} e_{it}$$

(8.15)

where $\phi_{1i}$ 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 preference across households. In particular, $\phi_{2it} = \exp(\tilde{\phi}_{2i} x_{it} + v_i)$ and $\phi_{3it} = \exp(\tilde{\phi}_{3i} x_{it} + t_i)$, where $v_i \sim N(0, \sigma_v^2)$ and $t_i \sim N(0, \sigma_t^2)$. 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.

Note that all parameters are index by $t$; age-specific parameters associated with children disutility cost follows from the existing work on child development. However, it is less common for parental time’s disutility cost to vary across children’s age. The age-specific parameter imply that parental time is a different consumption good (bad) depending on the age of the child.

The household has time constraints for both type 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 $\Omega_{it} = \{\theta_{it}, C_{it}, x_{it}, z_{it}\}$. The vector $z_{it}$ include demographic characteristics that influence the total factor productivity of the production function of cognitive skills. The vectors $z_{it}$ 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 household has rational expectations and forms expectation over three variables: future classroom inputs, skill shocks and future household income shocks. I assume classroom process is:

$$\log C_{it} = \kappa_t' x_{it} + \Delta_{it},$$  \hspace{1cm} (8.16)

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

$$\log y_{it} = \overline{y}_t + \rho_t \log y_{it} + \omega_{it},$$  \hspace{1cm} (8.17)

where $\{\overline{y}_t, \rho_t\}_t$ are parameters and $\omega_{it} \sim N(0, \sigma_{\omega,t}^2)$. The timing of the model is as follows: First, $v_{it-1}, \Delta_{it}$ and $\omega_{it}$ are realized at the beginning of the grade $t$, and the household learns

---

For example, Del Boca et al. (2019) assign age-dependent discount rate for children following the work by developmental psychologists that shows that the capacity to delay gratification changes significantly as 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)
Ω_{it}$. Next, the household makes the time investment decisions without knowing $v_{it}$.

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

$$V_T(\Omega_{i,T}) = \phi_4\frac{\theta_{1T}^{1-\phi_{1T}} - 1}{1 - \phi_{1T}}$$  \hspace{1cm} (8.18)

where $\phi_{4t} > 0$. The terminal value can be thought as the initial condition of a new problem.

### 8.2 Model estimation procedure

I implement a two-step estimation procedure to reduce computation 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 6.1 and 6.2, respectively. I estimate the household income and the classroom processes define in equation (8.17) and (8.16) with 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 $\Sigma$ and initial conditions, I simulate the households’ choices and skills across school grades.\textsuperscript{68} I then compute the analogous targeted moments in the simulated data, denoted by $M_S(\Sigma)$, and the estimator is defined as:

$$\hat{\Sigma} \equiv \arg\min_{\Sigma} (M - M_S(\Sigma))^\prime \times W \times (M - M_S(\Sigma)),$$ \hspace{1cm} (8.19)

where $W$ is the 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 complemented with (ii) first- and second-order moments for the conditional distribution of skills and time investments.

### 8.3 Model estimates

The estimates of the measurement system and skill technology are presented and described in Sections 7.1 and 7.2, respectively. Additionally, Table 13 have the estimates of the variance

\textsuperscript{68}The Appendix D describes the solution of the model.
of the skill technology shock at each school grade. Table 14 and 15 present the estimates of the household income and classroom processes. Finally, the estimates of the indirect inference estimator of the second step are in Table 16. 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.

8.4 Model fit

Figure 11 presents graphs that compare the values of the data moments with moments from the simulated data. There are four graphs, the top-left panel shows moments of the auxiliary model given by the policy functions of parental time and child effort and expected value and variance of the conditional distribution of skills. The top-right panel expands the graphs for the moments of the conditional distribution of skills, while the bottom panels expand the graphs for the moments of the parental and child time investment functions’ coefficients. Additionally, Figure 12 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 reasonable well the standard deviation of the skill, and somewhat larger standard deviation for time investment.

9 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 skills accumulation paths of children. For example, additional resources at grade 10 have a direct effect on students’ cognitive 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 weights between the direct effect of resources and the effects through self-productivity of skills. On top of this, households respond differently across grades, which has additional implications on the trade off.

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 framework, school resources are an abstract object given by class inputs’ effect on cognitive skills. 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 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 set of students in my sample, I observe if they enrolled in college the subsequent year of high school graduation. 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 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.

### 9.1 Policy-myopic households

In this exercise, I consider fourth graders in the sample and I simulate their classroom effects at each grade using equation (8.16). 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 13 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.\(^{70}\) The white bars show the distribution under the baseline allocation and the colored bars plot the distribution resulting from the allocation that maximizes the weighted skills index. For

---

\(^{70}\)This is similar to the exercise implemented in Cunha et al. (2010) regarding 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.
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 balance budget policy.

In addition, I perform the optimization setting as the objective function the terminal skills of students. Figure 14 shows the distribution of the difference between the classroom effects in tenth and fourth grades with terminal skills as the objective function. In contrast with the case with weighted skills index as the objective function, the distributions of the difference between late and early school investments shift toward the right—i.e., investing relatively more in 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 13.

To characterize the optimal allocation of students of different backgrounds, I split the distribution by demographic characteristics of students. The top panels of Figures 15 and 16 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 with a mothers with more than high school education have a (optimum) positive difference while this share is 20 percent for students with mothers with high school or less education. This is largely driven by higher self-productivity of skills for students from more affluent backgrounds.\(^{71}\)

### 9.2 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 assumption on the skill formation technology and the classroom process, a student has expected classroom inputs at each grade, conditional on \(x_{i0}\), given

\(^{71}\)See Figures S2 and S3, 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.
\[ R_{it} = \exp \left[ \int \cdots \int (\kappa_t x_{it} + \Delta_{it}) \times f_{\Delta,t}(\Delta_{it}) \times d\Delta_{it} \times \prod_{t' \leq t} f_{x_{it'}|x_{it'-1}}(x_{it'} | x_{it'-1}) \times dx_{it'} \right], \tag{9.20} \]

where \( f_{\Delta,t} \) is a density function of a normal distribution with mean zero and variance \( \sigma_{\Delta,t}^2 \) and \( f_{x_{it'}|x_{it'-1}}(x_{it'} | x_{it'-1}) \) is the transition function of \( x_{it} \). Thus, we can choose a policy that allocates optimally those resources across grades. A concern is that schools might face limitations to “move” 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 \textit{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:

\[ E[C_{it}] = 0.7 \times R_{it} + s_{it} \sum_{t'} 0.3 \times R_{it'}. \tag{9.21} \]

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 \( s^*_i \equiv \{s^*_it\}_t \) are defined as:

\[ s^*_i \equiv \arg\max_{s_i} \sum_t w_t \times \log \theta_{it} \quad \text{subject to} \quad \sum_t s_{it} = 1, \tag{9.22} \]

where the weights \( w_t \) are the coefficients of the regression of college attendance on skill measures.

\textit{Figure 17} shows the average (across students) optimal shares of transferable resources

---

72 The law of motion of classroom effects is \( \log C_{it} = \kappa_t x_{it} + \Delta_{it} \). Before attending grade \( t \), there is uncertainty about the realization of the classroom shocks \( \Delta_{it} \) and the income process shock \( \omega_{it} \) (note that \( x_{it} \) includes household income).
assigned at each school grade. On average, the shares are decreasing across grades; at fourth grade is around 47 percent and it decreases monotonously to 12 percent in grade 10. Figure 18 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 assign a larger share of transferable resources at lower grades.

Figure 19 shows the difference in average cognitive skills, classroom effects and time investment between the optimal and baseline resource allocations. The difference in skills follows a inverse-U shape across grades and they represent an increase of the weighted skills index of 0.2 SD of skills. 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 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 20 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 almost) achieved.

In addition, the bottom panel of Figure 20 shows the policy impact on the average weighted skills index for the top and bottom 20 percent of the household income distribution.

Figure S6 presents the distribution of the shares by school grade.
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 suggests that students from less affluent backgrounds benefit relatively more from the optimal allocation.

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 (9.22) for three households’ types: 1) forward-looking; 2) policy-myopic; and 3) no response. Figure 21 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 allocation resources at this grade. 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 and so respond differently.

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 households do not respond. I then calculate the impact of this allocation in 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 are designed internalizing the behavior of households. The difference between the outcomes inform about the importance of the behavioral response in the characterization of the optimal allocation.

In Figure 23, I plot the average difference in the weighted skills index between optimal and baseline allocation. Each bar assumes 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 optimal and baseline allocation assuming households do not alter their behavior.
not respond and the allocation is optimal for this household type. Under this scenario, the weighted skills index improves, on average, in 0.09 SD. This allocation would be the one we would characterize as optimal if we do 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 23—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 is given 0.17 and 0.20 SD for policy-myopic and forward-looking types, respectively—third (green) and last (dark blue) bars from the left of Figure 23. Not considering behavior in the characterization of optimal allocation of school resources implies considerably smaller average improvements in the weighed skills index, between 25 and 65 percent lower depending on the type of household behavior.75

10 Conclusions

In this paper, 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.76 Simulations of policies that remove the lowest-performing teachers (e.g., Hanushek, 2011; Goldhaber and Theobald, 2013; Chetty et al., 2014b) and analysis of

---

75Figure 24 shows the average difference between optimal and baseline allocation in skills at each grade associated with the scenarios presented in Figure 23.

76See Jackson et al. (2014) for a review of the literature on teacher-related policies.
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). Behrman et al. (2015) study teacher-incentive schemes and find no effect on academic achievement, unless the 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 of schemes that include a parent-incentive component, in addition to students’ and teachers’ components. 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.

Lastly, 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 in 0.20 SD with respect to the baseline allocation. On average, it is optimal for schools to invest relatively more on lower grades. Considering household behavioral responses plays a key role in the design of policies. The characterization of optimal allocation under the assumption of no behavioral response implies substantially smaller improvements if household do respond to school inputs; between 20 and 65 percent lower depending on the assumptions regarding the household’s expectations process.

77As 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.

78Levitt et al. (2016) study an incentive-based program with students and parents as beneficiaries, but it did not include a teacher-incentive component.

79The government introduced the policy in 2017 and it is gradually implemented until 2023. See Tincani (2020) for a detail description of the policy and an ex ante evaluation of the Chilean merit-based teaching reform.
References


Table 1: SIMCE data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

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 3.1.
Table 2: Descriptive statistics - Students

<table>
<thead>
<tr>
<th>School grade</th>
<th>Fourth</th>
<th>Sixth</th>
<th>Eighth</th>
<th>Tenth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Female student (%)</td>
<td>50.6</td>
<td>51.5</td>
<td>50.5</td>
<td>49.8</td>
</tr>
<tr>
<td>Mom answered (%)</td>
<td>79.8</td>
<td>81.5</td>
<td>80.4</td>
<td>78.9</td>
</tr>
<tr>
<td>Age parent (years)</td>
<td>38.3</td>
<td>7.3</td>
<td>40.1</td>
<td>7.2</td>
</tr>
<tr>
<td>Father &lt;HS (%)</td>
<td>28.3</td>
<td>29.2</td>
<td>32.6</td>
<td>30.8</td>
</tr>
<tr>
<td>Father HS (%)</td>
<td>36.4</td>
<td>36.6</td>
<td>35.5</td>
<td>36.4</td>
</tr>
<tr>
<td>Father &gt;HS (%)</td>
<td>35.2</td>
<td>34.2</td>
<td>31.9</td>
<td>32.8</td>
</tr>
<tr>
<td>Mother &lt;HS (%)</td>
<td>25.1</td>
<td>26.4</td>
<td>30.0</td>
<td>27.8</td>
</tr>
<tr>
<td>Mother HS (%)</td>
<td>38.4</td>
<td>36.6</td>
<td>38.0</td>
<td>39.3</td>
</tr>
<tr>
<td>Mother &gt;HS (%)</td>
<td>36.5</td>
<td>35.0</td>
<td>32.0</td>
<td>33.0</td>
</tr>
<tr>
<td>Monthly household income (ths. CLP)</td>
<td>617.2</td>
<td>579.9</td>
<td>608.5</td>
<td>575.7</td>
</tr>
<tr>
<td>Class size</td>
<td>33.4</td>
<td>7.6</td>
<td>34.7</td>
<td>7.4</td>
</tr>
<tr>
<td>No. teachers</td>
<td>5.8</td>
<td>2.2</td>
<td>8.6</td>
<td>1.4</td>
</tr>
<tr>
<td>No. subjects</td>
<td>9.8</td>
<td>0.8</td>
<td>10.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Classrooms</td>
<td>19,370</td>
<td>596,617</td>
<td>24,568</td>
<td>457,782</td>
</tr>
<tr>
<td>Students</td>
<td>407,720</td>
<td>596,617</td>
<td>457,782</td>
<td>336,470</td>
</tr>
</tbody>
</table>

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. Back to Section 3.2.
### Table 3:
Descriptive statistics - Teachers

<table>
<thead>
<tr>
<th>School grade</th>
<th>Fourth</th>
<th>Sixth</th>
<th>Eighth</th>
<th>Tenth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Female (%)</td>
<td>83.6</td>
<td>66.1</td>
<td>59.4</td>
<td>52.1</td>
</tr>
<tr>
<td>Age (years)</td>
<td>41.1</td>
<td>11.1</td>
<td>40.3</td>
<td>11.9</td>
</tr>
<tr>
<td>Education degree (%)</td>
<td>97.2</td>
<td>94.7</td>
<td>93.9</td>
<td>89.8</td>
</tr>
<tr>
<td>Other degree (%)</td>
<td>2.8</td>
<td>5.3</td>
<td>6.1</td>
<td>10.2</td>
</tr>
<tr>
<td>Teaching experience (years)</td>
<td>14.1</td>
<td>11.4</td>
<td>12.7</td>
<td>12.1</td>
</tr>
<tr>
<td>Tenure at school (years)</td>
<td>9.2</td>
<td>9.0</td>
<td>7.6</td>
<td>9.3</td>
</tr>
</tbody>
</table>

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 and additionally they teach at least two classrooms in different calendar years. Back to Section 3.2.

### Table 4:
Descriptive statistics - Parental and child time investment (weekly hours)

<table>
<thead>
<tr>
<th>Child’s age</th>
<th>School grade</th>
<th>Parental time</th>
<th>Child time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>10</td>
<td>4th</td>
<td>14.5</td>
<td>13.1</td>
</tr>
<tr>
<td>12</td>
<td>6th</td>
<td>13.0</td>
<td>12.8</td>
</tr>
<tr>
<td>14</td>
<td>8th</td>
<td>9.6</td>
<td>12.1</td>
</tr>
<tr>
<td>16</td>
<td>10th</td>
<td>6.5</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Note: Calculations are based on the Chilean Time Use Survey 2015. The survey reports hours spent in activities in the last week and weekend day. Parental time refers to hours parents spend in activities with their children and child time investment 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 3.2 or section 7.3.
Table 5: Examples of time investment questions in the SIMCE data

<table>
<thead>
<tr>
<th>Measure</th>
<th>Parent or child</th>
<th>Question and answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1 | C | My parents help me with my homework.  
| 2 | C | My parents help me study.  
| 3 | C | My parents explain me what I do not understand.  
| 4 | C | My parents are willing to help me when I have problems with a subject or homework.  
| 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.  
| 5 | C | I am a person who strives to learn.  
| 6 | C | I strive to get good grades.  

Note: These are examples of time investment questions reported in the parent and student questionnaires in the SIMCE administrative data. Back to Section 3.3.
Table 6: Measurement system of skills: WISC-V cognitive test

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

Source: Estimates using Chilean Wechsler Intelligence Scale for Children (WISC-V) survey.
Note: Test scores are assumed to be arbitrary scaled measures of skills $\theta_{it}$. 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 $\varepsilon_{itm}$ is measurement error and $\mu_m^A$ and $\lambda_m^A$ 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 (5.1) for additional details. Back to Section 7.1.
Table 7:
Measurement system of skills: administrative data tests

<table>
<thead>
<tr>
<th></th>
<th>Second grade</th>
<th>Fourth grade</th>
<th>Sixth grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_{nt}$</td>
<td>$\lambda_{nt}$</td>
<td>Signal share</td>
</tr>
<tr>
<td>Math SIMCE</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Language SIMCE</td>
<td>253.1</td>
<td>29.2</td>
<td>0.35</td>
</tr>
<tr>
<td>Natural sciences SIMCE</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Social sciences SIMCE</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Math. grade</td>
<td>5.6</td>
<td>0.7</td>
<td>0.69</td>
</tr>
<tr>
<td>Language grade</td>
<td>5.5</td>
<td>0.8</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Eighth grade</th>
<th>Tenth grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_{nt}$</td>
<td>$\lambda_{nt}$</td>
</tr>
<tr>
<td>Math SIMCE</td>
<td>165.3</td>
<td>32.0</td>
</tr>
<tr>
<td>Language SIMCE</td>
<td>146.8</td>
<td>32.9</td>
</tr>
<tr>
<td>Natural sciences SIMCE</td>
<td>176.4</td>
<td>29.7</td>
</tr>
<tr>
<td>Social sciences SIMCE</td>
<td>170.6</td>
<td>30.2</td>
</tr>
<tr>
<td>Math grade</td>
<td>3.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Language grade</td>
<td>4.0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Source: Estimates using SIMCE and SIGE administrative databases.
Note: Test scores are assumed to be arbitrary scaled measures of skills $\theta_{it}$. 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 $\varepsilon_{itm}$ is measurement error and $\mu_{mt}$ and $\lambda_{mt}$ 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 (5.4) for additional details. Back to Section 7.1.
Table 8: 
Skill formation technology

<table>
<thead>
<tr>
<th></th>
<th>Fourth (1)</th>
<th>Sixth (2)</th>
<th>Eighth (3)</th>
<th>Tenth (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $\theta_{it}$ (Skill)</td>
<td>0.795</td>
<td>0.495</td>
<td>0.588</td>
<td>0.493</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.008)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td></td>
<td>[0.761;0.829]</td>
<td>[0.480;0.511]</td>
<td>[0.569;0.61]</td>
<td>[0.471;0.519]</td>
</tr>
<tr>
<td>$\log^2 \theta_{it}$</td>
<td>0.066</td>
<td>0.063</td>
<td>0.034</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>[0.06;0.071]</td>
<td>[0.06;0.066]</td>
<td>[0.031;0.038]</td>
<td>[0.000;0.007]</td>
</tr>
<tr>
<td>$h_{it}$ (parental time - daily hours)</td>
<td>0.442</td>
<td>0.104</td>
<td>0.278</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.020)</td>
<td>(0.037)</td>
<td>(0.031)</td>
</tr>
<tr>
<td></td>
<td>[0.340;0.547]</td>
<td>[0.069;0.144]</td>
<td>[0.203;0.349]</td>
<td>[0.082;0.206]</td>
</tr>
<tr>
<td>$h_{it}^2$</td>
<td>-0.056</td>
<td>-0.035</td>
<td>-0.059</td>
<td>-0.048</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.009)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td>[-0.074;-0.039]</td>
<td>[-0.045;-0.025]</td>
<td>[-0.077;-0.040]</td>
<td>[-0.064;-0.032]</td>
</tr>
<tr>
<td>$e_{it}$ (child time - daily hours)</td>
<td>-1.787</td>
<td>0.462</td>
<td>0.096</td>
<td>-0.095</td>
</tr>
<tr>
<td></td>
<td>(1.074)</td>
<td>(0.043)</td>
<td>(0.043)</td>
<td>(0.046)</td>
</tr>
<tr>
<td></td>
<td>[-3.968;0.361]</td>
<td>[0.375;0.544]</td>
<td>[0.003;0.18]</td>
<td>[-0.183;-0.006]</td>
</tr>
<tr>
<td>$e_{it}^2$</td>
<td>8.106</td>
<td>-1.153</td>
<td>-1.132</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>(3.103)</td>
<td>(0.017)</td>
<td>(0.029)</td>
<td>(0.019)</td>
</tr>
<tr>
<td></td>
<td>[1.968;14.726]</td>
<td>[-0.185;-0.117]</td>
<td>[-0.187;-0.077]</td>
<td>[0.001;0.075]</td>
</tr>
<tr>
<td>log $C_{it}$ (classroom effects)</td>
<td>0.577</td>
<td>0.479</td>
<td>0.541</td>
<td>0.422</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.008)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td></td>
<td>[0.542;0.608]</td>
<td>[0.461;0.494]</td>
<td>[0.516;0.565]</td>
<td>[0.394;0.444]</td>
</tr>
</tbody>
</table>

*(table continues in next page)*
Table 8: — Continued
Skill formation technology

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

$N$ = 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 7.2.
<table>
<thead>
<tr>
<th></th>
<th>Fourth (1)</th>
<th>Sixth (2)</th>
<th>Eighth (3)</th>
<th>Tenth (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log \theta_{it} ) (Skill)</td>
<td>0.815</td>
<td>0.487</td>
<td>0.497</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.009)</td>
<td>(0.033)</td>
<td>(0.035)</td>
</tr>
<tr>
<td></td>
<td>[0.765;0.858]</td>
<td>[0.467;0.504]</td>
<td>[0.476;0.508]</td>
<td>[0.485;0.533]</td>
</tr>
<tr>
<td>( \log^2 \theta_{it} )</td>
<td>0.062</td>
<td>0.059</td>
<td>0.068</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>[0.053;0.071]</td>
<td>[0.055;0.062]</td>
<td>[0.065;0.073]</td>
<td>[-0.003;0.003]</td>
</tr>
<tr>
<td>( h_{it} ) (parental time - daily hours)</td>
<td>0.073</td>
<td>0.119</td>
<td>0.035</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.020)</td>
<td>(0.033)</td>
<td>(0.027)</td>
</tr>
<tr>
<td></td>
<td>[-0.018;0.181]</td>
<td>[0.073;0.152]</td>
<td>[-0.027;0.100]</td>
<td>[-0.06;0.044]</td>
</tr>
<tr>
<td>( h_{it}^2 )</td>
<td>-0.009</td>
<td>-0.030</td>
<td>-0.016</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.004)</td>
<td>(0.008)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>[-0.029;0.014]</td>
<td>[-0.037;0.021]</td>
<td>[-0.031;0]</td>
<td>[-0.009;0.013]</td>
</tr>
<tr>
<td>( e_{it} ) (child time - daily hours)</td>
<td>2.200</td>
<td>0.576</td>
<td>0.238</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(1.427)</td>
<td>(0.041)</td>
<td>(0.059)</td>
<td>(0.042)</td>
</tr>
<tr>
<td></td>
<td>[-0.507;4.984]</td>
<td>[0.516;0.682]</td>
<td>[0.110;0.346]</td>
<td>[-0.049;0.114]</td>
</tr>
<tr>
<td>( e_{it}^2 )</td>
<td>-5.263</td>
<td>-0.204</td>
<td>-0.100</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>(4.277)</td>
<td>(0.020)</td>
<td>(0.043)</td>
<td>(0.020)</td>
</tr>
<tr>
<td></td>
<td>[-13.523;2.677]</td>
<td>[-0.254;0.174]</td>
<td>[-0.187;0.013]</td>
<td>[0.004;0.086]</td>
</tr>
<tr>
<td>( \log T_{it} ) (teachers effects)</td>
<td>0.437</td>
<td>0.375</td>
<td>0.493</td>
<td>0.350</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.012)</td>
<td>(0.039)</td>
<td>(0.029)</td>
</tr>
<tr>
<td></td>
<td>[0.388;0.476]</td>
<td>[0.345;0.391]</td>
<td>[0.456;0.545]</td>
<td>[0.313;0.385]</td>
</tr>
</tbody>
</table>

*(table continues in next page)*
Table 9: — Continued
Skill formation technology - specification with within classroom inputs

<table>
<thead>
<tr>
<th>Skill formation technology</th>
<th>School grade</th>
<th>Fourth (1)</th>
<th>Sixth (2)</th>
<th>Eighth (3)</th>
<th>Tenth (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $\theta_{it} \times \log T_{it}$</td>
<td></td>
<td>-0.040</td>
<td>0.008</td>
<td>-0.037</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.006)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.050;-0.031]</td>
<td>[0.001;0.013]</td>
<td>[-0.052;-0.031]</td>
<td>[-0.026;-0.011]</td>
</tr>
<tr>
<td>$h_{it} \times \log T_{it}$</td>
<td></td>
<td>-0.006</td>
<td>-0.023</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.021;0.007]</td>
<td>[-0.038;-0.008]</td>
<td>[-0.008;0.008]</td>
<td>[-0.013;0.011]</td>
</tr>
<tr>
<td>$e_{it} \times \log T_{it}$</td>
<td></td>
<td>0.165</td>
<td>0.061</td>
<td>0.006</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.119)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.048;0.436]</td>
<td>[0.039;0.086]</td>
<td>[-0.016;0.023]</td>
<td>[-0.014;0.018]</td>
</tr>
<tr>
<td>$h_{it} \times e_{it}$</td>
<td></td>
<td>0.151</td>
<td>-0.067</td>
<td>0.070</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.233)</td>
<td>(0.024)</td>
<td>(0.021)</td>
<td>(0.021)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.271;0.636]</td>
<td>[-0.108;-0.019]</td>
<td>[0.024;0.106]</td>
<td>[-0.062;0.020]</td>
</tr>
<tr>
<td>$h_{it} \times \log \theta_{it}$</td>
<td></td>
<td>-0.025</td>
<td>0.063</td>
<td>-0.007</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[-0.037;-0.010]</td>
<td>[0.049;0.075]</td>
<td>[-0.013;0.001]</td>
<td>[0.001;0.022]</td>
</tr>
<tr>
<td>$e_{it} \times \log \theta_{it}$</td>
<td></td>
<td>0.267</td>
<td>0.070</td>
<td>0.055</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.133)</td>
<td>(0.008)</td>
<td>(0.011)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.007;0.525]</td>
<td>[0.055;0.087]</td>
<td>[0.034;0.077]</td>
<td>[0.032;0.060]</td>
</tr>
</tbody>
</table>

$N$ 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 7.2.
Table 10:
Validation Test: Selection on observables — Household income

<table>
<thead>
<tr>
<th>School grade</th>
<th>Fourth (1)</th>
<th>Sixth (2)</th>
<th>Eighth (3)</th>
<th>Tenth (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $\theta_{it+1}$ (skills)</td>
<td>0.1352 (0.0043)</td>
<td>0.1531 (0.0045)</td>
<td>0.1824 (0.0055)</td>
<td>0.1266 (0.0042)</td>
</tr>
<tr>
<td>[0.1271;0.1436]</td>
<td>[0.1445;0.1621]</td>
<td>[0.1721;0.1937]</td>
<td>[0.1187;0.1353]</td>
<td></td>
</tr>
<tr>
<td>$\hat{\varphi}_{it}$ (residual skill tech.)</td>
<td>0.0033 (0.0004)</td>
<td>0.0028 (0.0002)</td>
<td>0.0037 (0.0003)</td>
<td>0.0007 (0.0002)</td>
</tr>
<tr>
<td>[0.0025;0.004]</td>
<td>[0.0023;0.0032]</td>
<td>[0.0031;0.0041]</td>
<td>[0.0001;0.0011]</td>
<td></td>
</tr>
</tbody>
</table>

| $N$ | 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 $\hat{\varphi}_{it}$ represents the residual of the estimated technology from the specification in Table 8. The first row shows the coefficients of regressing skills on household income, measured in $3000$ dollars US in 2018 values (0.3 SD of the household annual income distribution). Instead, the second row shows the coefficients of regressing the technology’s residuals on household income. Back to Section 7.2.

Table 11:
Coefficient of regression of classroom effects on household income

<table>
<thead>
<tr>
<th>School grade</th>
<th>Fourth (1)</th>
<th>Sixth (2)</th>
<th>Eighth (3)</th>
<th>Tenth (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No school FE</td>
<td>0.0493 (0.0039)</td>
<td>0.0704 (0.0028)</td>
<td>0.0906 (0.0037)</td>
<td>0.0896 (0.0007)</td>
</tr>
<tr>
<td>[0.0418;0.0573]</td>
<td>[0.0648;0.0764]</td>
<td>[0.0812;0.0975]</td>
<td>[0.0840;0.0967]</td>
<td></td>
</tr>
<tr>
<td>School FE</td>
<td>0.0080 (0.0007)</td>
<td>-0.0035 (0.0008)</td>
<td>0.0039 (0.0011)</td>
<td>0.0083 (0.0008)</td>
</tr>
<tr>
<td>[0.0067;0.0093]</td>
<td>[-0.0052;-0.0021]</td>
<td>[0.0027;0.0067]</td>
<td>[0.0068;0.0098]</td>
<td></td>
</tr>
</tbody>
</table>

| $N$ | 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 of regressing classroom effects on household income, measured in $3000$ 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 8. Back to Section 7.2.
Table 12:
Parental and child time (weekly hours) responses to classroom and teacher effects

<table>
<thead>
<tr>
<th></th>
<th>Parental time</th>
<th>Child time</th>
<th>Parental time</th>
<th>Child time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>log $E_{it}$ = log $C_{it}$ (classroom effects)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental time</td>
<td>(1)</td>
<td>(2)</td>
<td>Parental time</td>
<td>(3)</td>
</tr>
<tr>
<td>log $E_{it}$</td>
<td>-1.353</td>
<td>0.240</td>
<td>-0.557</td>
<td>0.172</td>
</tr>
<tr>
<td>(0.129)</td>
<td>(0.081)</td>
<td>(0.124)</td>
<td>(0.054)</td>
<td></td>
</tr>
<tr>
<td>[-1.556; -1.143]</td>
<td>[0.170; 0.306]</td>
<td>[-0.753; -0.293]</td>
<td>[0.101; 0.282]</td>
<td></td>
</tr>
<tr>
<td>log $E_{it} \times 1 {\text{Grade} = 6}$</td>
<td>0.723</td>
<td>0.137</td>
<td>-0.073</td>
<td>0.091</td>
</tr>
<tr>
<td>(0.112)</td>
<td>(0.043)</td>
<td>(0.107)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>[0.508; 0.942]</td>
<td>[0.058; 0.182]</td>
<td>[-0.290; 0.136]</td>
<td>[-0.018; 0.125]</td>
<td></td>
</tr>
<tr>
<td>log $E_{it} \times 1 {\text{Grade} = 8}$</td>
<td>1.112</td>
<td>0.049</td>
<td>0.473</td>
<td>0.055</td>
</tr>
<tr>
<td>(0.109)</td>
<td>(0.046)</td>
<td>(0.117)</td>
<td>(0.046)</td>
<td></td>
</tr>
<tr>
<td>[0.952; 1.370]</td>
<td>[-0.030; 0.155]</td>
<td>[0.249; 0.698]</td>
<td>[-0.028; 0.146]</td>
<td></td>
</tr>
<tr>
<td>log $E_{it} \times 1 {\text{Grade} = 10}$</td>
<td>1.932</td>
<td>0.168</td>
<td>0.806</td>
<td>-0.269</td>
</tr>
<tr>
<td>(0.148)</td>
<td>(0.083)</td>
<td>(0.142)</td>
<td>(0.063)</td>
<td></td>
</tr>
<tr>
<td>[1.740; 2.215]</td>
<td>[0.127; 0.266]</td>
<td>[0.482; 1.008]</td>
<td>[-0.418; -0.185]</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1,079,935</td>
<td>1,079,935</td>
<td>555,415</td>
<td>555,415</td>
</tr>
</tbody>
</table>

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 7.3.
Table 13: Standard deviation of the skill technology shock

<table>
<thead>
<tr>
<th>School grades</th>
<th>4th grade</th>
<th>6th grade</th>
<th>8th grade</th>
<th>10th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{v,t}$</td>
<td>1.209</td>
<td>0.869</td>
<td>0.982</td>
<td>0.711</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>[1.202;1.222]</td>
<td>[0.863;0.874]</td>
<td>[0.977;0.986]</td>
<td>[0.702;0.722]</td>
</tr>
</tbody>
</table>

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 8. Back to Section 8.3.

Table 14: Household income process

<table>
<thead>
<tr>
<th>School grades</th>
<th>4th grade</th>
<th>6th grade</th>
<th>8th grade</th>
<th>10th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{y}_t$</td>
<td>-</td>
<td>3.434</td>
<td>3.365</td>
<td>3.383</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.024)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[3.397;3.498]</td>
<td>[3.336;3.43]</td>
<td>[3.335;3.433]</td>
<td></td>
</tr>
<tr>
<td>$\rho_t^y$</td>
<td>-</td>
<td>0.743</td>
<td>0.749</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.738;0.746]</td>
<td>[0.744;0.751]</td>
<td>[0.744;0.751]</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\omega,t}$</td>
<td>-</td>
<td>0.269</td>
<td>0.268</td>
<td>0.285</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.266;0.273]</td>
<td>[0.266;0.273]</td>
<td>[0.282;0.289]</td>
<td></td>
</tr>
</tbody>
</table>

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 (8.17). Back to Section 8.3.
Table 15: Classroom process

<table>
<thead>
<tr>
<th></th>
<th>4th grade</th>
<th>6th grade</th>
<th>8th grade</th>
<th>10th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.260</td>
<td>-0.525</td>
<td>-0.610</td>
<td>-0.804</td>
</tr>
<tr>
<td></td>
<td>(0.0270)</td>
<td>(0.0213)</td>
<td>(0.0272)</td>
<td>(0.0295)</td>
</tr>
<tr>
<td></td>
<td>[-0.315;-0.208]</td>
<td>[-0.57;-0.485]</td>
<td>[-0.664;-0.56]</td>
<td>[-0.866;-0.752]</td>
</tr>
<tr>
<td>Montly HH income (ths. CLP)</td>
<td>0.250</td>
<td>0.258</td>
<td>0.355</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>(0.0185)</td>
<td>(0.0145)</td>
<td>(0.0140)</td>
<td>(0.0116)</td>
</tr>
<tr>
<td></td>
<td>[0.216;0.288]</td>
<td>[0.229;0.285]</td>
<td>[0.33;0.383]</td>
<td>[0.312;0.357]</td>
</tr>
<tr>
<td>Father HS</td>
<td>0.019</td>
<td>0.084</td>
<td>0.099</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>(0.0083)</td>
<td>(0.0078)</td>
<td>(0.0088)</td>
<td>(0.0091)</td>
</tr>
<tr>
<td></td>
<td>[0.003;0.035]</td>
<td>[0.069;0.1]</td>
<td>[0.083;0.115]</td>
<td>[0.12;0.156]</td>
</tr>
<tr>
<td>Father &gt;HS</td>
<td>0.029</td>
<td>0.176</td>
<td>0.177</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>(0.0120)</td>
<td>(0.0102)</td>
<td>(0.0120)</td>
<td>(0.0129)</td>
</tr>
<tr>
<td></td>
<td>[0.006;0.053]</td>
<td>[0.157;0.197]</td>
<td>[0.153;0.201]</td>
<td>[0.206;0.256]</td>
</tr>
<tr>
<td>Mother HS</td>
<td>0.055</td>
<td>0.143</td>
<td>0.161</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>(0.0088)</td>
<td>(0.0075)</td>
<td>(0.0079)</td>
<td>(0.0091)</td>
</tr>
<tr>
<td></td>
<td>[0.038;0.073]</td>
<td>[0.127;0.157]</td>
<td>[0.146;0.176]</td>
<td>[0.185;0.221]</td>
</tr>
<tr>
<td>Mother &gt;HS</td>
<td>0.066</td>
<td>0.217</td>
<td>0.239</td>
<td>0.304</td>
</tr>
<tr>
<td></td>
<td>(0.0114)</td>
<td>(0.0101)</td>
<td>(0.0106)</td>
<td>(0.0123)</td>
</tr>
<tr>
<td></td>
<td>[0.043;0.089]</td>
<td>[0.197;0.237]</td>
<td>[0.218;0.26]</td>
<td>[0.282;0.331]</td>
</tr>
<tr>
<td>Age parent</td>
<td>0.002</td>
<td>0.005</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td></td>
<td>[0.002;0.003]</td>
<td>[0.004;0.005]</td>
<td>[0.005;0.007]</td>
<td>[0.008;0.009]</td>
</tr>
<tr>
<td>( \sigma^2_\Delta )</td>
<td>0.926</td>
<td>0.901</td>
<td>0.875</td>
<td>0.864</td>
</tr>
<tr>
<td></td>
<td>(0.0073)</td>
<td>(0.0068)</td>
<td>(0.0097)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td></td>
<td>[0.909;0.939]</td>
<td>[0.887;0.912]</td>
<td>[0.855;0.893]</td>
<td>[0.843;0.881]</td>
</tr>
<tr>
<td>( N )</td>
<td>407,720</td>
<td>596,617</td>
<td>457,782</td>
<td>336,470</td>
</tr>
</tbody>
</table>

Note: Schools network-clustered bootstrapped standard errors and 95% confidence intervals in parentheses and brackets respectively. The classroom process is \( \log C_{it} = \kappa' x_{it} + \Delta_{it} \), where \( \Delta_{it} \sim N(0, \sigma^2_{\Delta, it}) \), is estimated with OLS estimator. For additional details see equation (8.16). 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 8.3.
Table 16: Estimates preference parameters

<table>
<thead>
<tr>
<th>School grade</th>
<th>Fourth</th>
<th>Sixth</th>
<th>Eighth</th>
<th>Tenth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>CRRA utility parameter on skills: $\phi_{1t}$</td>
<td>0.566 (0.0001)</td>
<td>0.053 (0.0001)</td>
<td>0.383 (0)</td>
<td>0.061 (0.0108)</td>
</tr>
<tr>
<td>Parent time disutility costs parameters: $\tilde{\phi}_{2it}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>4.579 (0.1081)</td>
<td>4.610 (0.0145)</td>
<td>2.336 (0.0549)</td>
<td>5.859 (0.0698)</td>
</tr>
<tr>
<td>HH income</td>
<td>-1.798 (0.0048)</td>
<td>-0.040 (0.0052)</td>
<td>-0.261 (0.0004)</td>
<td>-0.023 (0.0003)</td>
</tr>
<tr>
<td>Father HS</td>
<td>-2.228 (0.0079)</td>
<td>-0.279 (0.0002)</td>
<td>-0.003 (0.0061)</td>
<td>-1.781 (0.007)</td>
</tr>
<tr>
<td>Father &gt;HS</td>
<td>0.000 (0.0002)</td>
<td>-1.451 (0.0313)</td>
<td>-0.011 (0.0003)</td>
<td>-0.004 (0.0001)</td>
</tr>
<tr>
<td>Mother HS</td>
<td>-0.046 (0.0001)</td>
<td>-2.803 (0.0074)</td>
<td>-0.062 (0.0009)</td>
<td>-0.391 (0.0012)</td>
</tr>
<tr>
<td>Mother &gt;HS</td>
<td>-0.412 (0.0005)</td>
<td>-0.007 (0.0007)</td>
<td>-0.001 (0.0003)</td>
<td>-0.371 (0.0001)</td>
</tr>
<tr>
<td>Parent age</td>
<td>-1.414 (0.0041)</td>
<td>-2.218 (0.0006)</td>
<td>-2.215 (0.0065)</td>
<td>-1.332 (0.0036)</td>
</tr>
<tr>
<td>Student time disutility costs parameters: $\tilde{\phi}_{3it}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.936 (0.1383)</td>
<td>4.285 (0.0144)</td>
<td>4.913 (0.0112)</td>
<td>5.257 (0.1022)</td>
</tr>
<tr>
<td>HH income</td>
<td>1.779 (0.0035)</td>
<td>1.413 (0.0050)</td>
<td>1.934 (0.0028)</td>
<td>-2.550 (0.0057)</td>
</tr>
<tr>
<td>Father HS</td>
<td>1.765 (0.0020)</td>
<td>-1.495 (0.0015)</td>
<td>0.544 (0.0023)</td>
<td>0.620 (0.0045)</td>
</tr>
<tr>
<td>Father &gt;HS</td>
<td>1.373 (0.0016)</td>
<td>-1.535 (0.0041)</td>
<td>0.356 (0.0027)</td>
<td>1.987 (0.0035)</td>
</tr>
<tr>
<td>Mother HS</td>
<td>1.985 (0.0020)</td>
<td>-0.237 (0.0067)</td>
<td>-2.509 (0.0049)</td>
<td>1.114 (0.0038)</td>
</tr>
<tr>
<td>Mother &gt;HS</td>
<td>1.478 (0.0016)</td>
<td>-0.236 (0.0019)</td>
<td>-1.699 (0.0038)</td>
<td>1.762 (0.0037)</td>
</tr>
<tr>
<td>Parent age</td>
<td>1.991 (0.0047)</td>
<td>-0.569 (0.0002)</td>
<td>-1.323 (0.0023)</td>
<td>-2.230 (0.0055)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal period parameter: $\phi_4$</td>
<td>1.007 (0.0015)</td>
</tr>
<tr>
<td>Variance $\psi_i$: $\sigma^2_\psi$</td>
<td>0.025 (0.0001)</td>
</tr>
<tr>
<td>Variance $\psi_i$: $\sigma^2_i$</td>
<td>0.059 (0.0020)</td>
</tr>
</tbody>
</table>

Note: Schools network-clustered bootstrapped standard errors in parentheses. The preference parameters are specified as: CRRA parameter on skills $\phi_{1t}$, parental and student time disutility costs $\tilde{\phi}_{2it} = \exp(\tilde{\phi}_{2it}'x_{it} + \psi_i)$ and $\tilde{\phi}_{3it} = \exp(\tilde{\phi}_{3it}'x_{it} + \psi_i)$, respectively, where $\psi_i \sim N(0, \sigma^2_\psi)$ and $\psi_i \sim N(0, \sigma^2_i)$. Terminal period parameter $\phi_4$. For additional detail see Section 8.1. The vector $x_{it}$ 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 8.3.
Figures

Figure 1:
Ordered categorical questions of time investment and classroom value-added

Parental time

![Parental time graph]

Child effort

![Child effort graph]

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. Back to Section 3.3.
Figure 2:
Parental and child time investment distributions

Parental time
Child effort

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 3.2.

Figure 3:
Expected value and standard deviation of skills

Note: Estimated with the Chilean WISC-V cognitive development test survey. The estimation consist on replacing the moments of equation (5.3) with their sample analogs and the measurement system’s parameters presented in Table 6. Back to Section 7.1.
Figure 4:
Sample average marginal effects of skill formation inputs (in SD log skill 2nd grade)

Parental time (weekly hours)

Child time (weekly hours)

Classrooms (SD)

Current skills (SD 2nd grade)

Note: The values on these graphs show the average marginal effect calculated using the estimates from the specifications in Table 8. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student’s marginal effect using each input’s analogous equation (7.14) and calculate the average over the sample. Back to Section 7.2.
Figure 5:
Sample average marginal effects of skill formation inputs (in percentiles of the skill distribution)

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 8. 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 (7.14) and calculate the average over the sample. Back to Section 7.2.
Figure 6:
Sample average marginal effects of skill formation inputs (in SD log skill 2nd grade) specification with within classroom components

Parental time (weekly hours)

Child time (weekly hours)

Classrooms (SD)

Current skills (SD 2nd grade)

Note: The values on these graphs show the average marginal effect calculated using the estimates from the specifications in Table 9. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student’s marginal effect using each input’s analogous equation (7.14) and calculate the average over the sample. Back to Section 7.2.
Figure 7:
Sample average marginal effects of skill formation inputs (in percentiles of the skill distribution)
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 9. 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 (7.14) and calculate the average over the sample. Back to Section 7.2.
Figure 8:
Validation test: skill formation technology
Out-of-sample fit of skills by household income

Fourth grade

Sixth grade

Eighth grade

Tenth grade

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 7.2.
Figure 9:
Parental time responses (weekly hours) to reassignment from 25th to 75th percentile of classroom and teacher quality distributions

Classrooms

Teachers

Note: The values on the graphs are calculated using the estimates of Table 12. 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. Back to Section 7.3.

Figure 10:
Child time responses (weekly hours) to reassignment from 25th to 75th percentile of classroom and teacher quality distributions

Classrooms

Teachers

Note: The values on the graphs are calculated using the estimates of Table 12. 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. Back to Section 7.3.
Figure 11: 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 8.4.
Figure 12:  
Model fit of targeted moments  
Mean and SD of skills and parental and child time investment

### Skills

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Parental time

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Child time self-investment

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The bars labeled data and model show the mean and standard deviation (SD) estimated in the data and data simulated with the child development model. Back to Section 8.4.
Note: The white bars plot the distribution of the difference between the classroom effects in grades 10 and 4, that is 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 (8.16). Then, with the time investment functions (Table 12) and the skill technology (Table 8) I simulate the household choices and their skills for every possible assignment of the realizations of classroom effects \( \{ \log C_{it} \} \) 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 household respond as if resources in subsequent grades are the given by baseline allocation. See section 9 for additional details.
Figure 14:
Policy myopic households: Optimal allocation that maximize skills at grade 10 by mother education

Note: The white bars plot the distribution of the difference between the classroom effects in grades 10 and 4, that is 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 (8.16). Then, with the time investment functions (Table 12) and the skill technology (Table 8) 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 skills at grade 10 of each student. Note that using the investment functions for the simulations implicitly assumes that household respond as if resources in subsequent grades are the given by baseline allocation. See section 9 for additional details.
Figure 15:
Policy myopic households: Optimal allocation that maximize (weighted) skills by mother education

Note: The top panel shows the distributions by mother education of the difference between the classroom effect in grade 10 and 4 of the optimal allocation. The red (light color) and blue (dark color) bars are mother with high school or less and more than high school education, receptively. The bottom panel shows the associated cumulative distributions. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). Then, with the time investment functions (Table 12) and the skill technology (Table 8) 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 household respond as if resources in subsequent grades are the given by baseline allocation. See section 9 for additional details.
Figure 16:
Policy myopic households: Optimal allocation that maximize (weighted) skills by household income

Note: The top panel shows the distributions by household income quintiles of the difference between the classroom effect in grade 10 and 4 of the optimal allocation. The red (light color) and blue (dark color) bars are bottom and top 20 percent of the household income distribution, receptively. The bottom panel shows the associated cumulative distributions. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). Then, with the time investment functions (Table 12) and the skill technology (Table 8) 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 household respond as if resources in subsequent grades are the given by baseline allocation. See section 9 for additional details.
Figure 17:
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 (8.16). 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 9.2 for additional details.
Figure 18:
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 (8.16). 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 9.2 for additional details.
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. For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). 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 9.2 for additional details.
Figure 20:
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 (8.16). 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 9.2 for additional details.
Figure 21:
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 8). For policy myopic household I simulate their choices and skills with the time investment functions (Table 12) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 8.1). For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). 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 9.2 for additional details.
Figure 22:
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 8). For policy myopic household I simulate their choices and skills with the time investment functions (Table 12) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 8.1). For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). 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 9.2 for additional details.
Figure 23:
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 8). For policy myopic household I simulate their choices and skills with the time investment functions (Table 12) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 8.1). For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). 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 9.2 for additional details.
Figure 24:
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 8). For policy myopic household I simulate their choices and skills with the time investment functions (Table 12) and skill technology. For forward-looking households I simulate their choices and skills with the full child development model (see section 8.1). For students attending fourth grade in the sample I draw classroom effects for grades 4 to 10 using equation (8.16). 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 9.2 for additional details.
Appendices

A Data

A.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.

A.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 tests. 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.

B Time investments measurement

B.1 Results identification measurement system

The identification of the parameters of the response model follows San Martín et al. (2013). Proof that $\eta(\cdot)$ is increasing in $\beta$. 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 $s$.

$$ p_{s,s'} = \int F(\beta_s h - \alpha(\beta_s, p_s))F(\beta_{s'} h - \alpha(\beta_{s'}, p_{s'}))g(h)dh \equiv q(\beta_s, p_s, \beta_{s'}, p_{s'}) \quad (B.1) $$
Taking the derivative of equation (B.1) with respect to \( \beta_s \) and rearranging:

\[
\frac{\partial q(\beta_s, p_s, \beta_s', p_s')}{\partial \beta_s} = \int f(\beta_s h - \alpha(\beta_s, p_s))(h - \frac{\partial \alpha}{\partial \beta_s}(\beta_s, p_s))F(\beta_s h - \alpha(\beta_s', p_s'))g(h)dh \tag{B.2}
\]

Replace \( \alpha(\beta_s, p_s) \) in equation (5.7) and take the derivative with respect to \( \beta_1 \) and rearrange:

\[
\frac{\partial \alpha}{\partial \beta_s}(\beta_s, p_s) = \int h g_C(h; \beta_s, p_s)dh = E_{\beta_s, \alpha}[h] \tag{B.3}
\]

where

\[
g_C(h; \beta_s, p_s) = \frac{f(\beta_s h - \alpha(\beta_s, p_s))g(h_i)}{\int f(\beta_s h - \alpha(\beta_s, p_s))g(h)dh} \tag{B.4}
\]

Multiply and divide equation (B.2) by \( \int f(\beta_s h - \alpha(\beta_s, p_s))g(h)dh \), we can rewrite the equation as:

\[
\frac{\partial q(\beta_s, p_s, \beta_s', p_s')}{\partial \beta_s} = \int (h - E_{\beta_s, \alpha}[h])F(\beta_s h - \alpha(\beta_s', p_s'))g_C(h; \beta_s, p_s)dh \tag{B.5}
\]

Then, define:

\[
E_{\beta_s, \alpha}[F(\beta_s h - \alpha(\beta_s', p_s'))] = \int F(\beta_s h - \alpha(\beta_s', p_s'))g_C(h; \beta_s, p_s)dh \tag{B.6}
\]

Subtracting \( E_{\beta_s, p_s}[F(\beta_s h - \alpha(\beta_s', p_s'))] \int (h - E_{\beta_s, \alpha}[h])g_C(h; \beta_s, p_s)dh = 0 \) on left side of equation (B.5) and rearranging:

\[
\frac{\partial q(\beta_s, p_s, \beta_s', p_s')}{\partial \beta_s} = \\
\int (h - E_{\beta_s, \alpha}(\beta_s, p_s)[h])(F(\beta_s h - \alpha(\beta_s', p_s')) - E_{\beta_s, \alpha}(\beta_s, p_s)[F(\beta_s h - \alpha(\beta_s', p_s'))])g_C(h; \beta_s, p_s)dh \\
\times \int f(\beta_s h - \alpha(\beta_s, p_s))g(h_i)dh_i \\
= \text{Cov}_{\beta_s, \alpha}(\beta_s, p_s)(h, F[\beta_s h - \alpha(\beta_s', p_s')]) \times \int f(\beta_s h - \alpha(\beta_s, p_s))g(h)dh > 0 \tag{B.7}
\]

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 \( \beta_1 \).

\[
\frac{\partial r}{\partial \beta_1}(\beta_1, p_1, p_2, p_3, p_{1,2}, p_{1,3}) = \\
\frac{\partial q}{\partial \beta_2}(\beta_1, p_1, p_{1,2}, p_2, p_2, \bar{q}(\beta_1, p_1, p_{1,3}, p_3, p_3)) \times \frac{\partial \bar{q}}{\partial \beta_1}(\beta_1, p_1, p_{1,2}, p_2) \\
+ \frac{\partial q}{\partial \beta_3}(\beta_1, p_1, p_{1,2}, p_2, p_2, \bar{q}(\beta_1, p_1, p_{1,3}, p_3, p_3)) \times \frac{\partial \bar{q}}{\partial \beta_1}(\beta_1, p_1, p_{1,3}, p_3)
\]

(B.8)

Since \( \frac{\partial q}{\partial \beta_s} > 0 \) for all \( s \) and given that:

\[
q(\beta_1, p_1, \bar{q}(\beta_1, p_1, p_{1,2}, p_2), p_2) = p_{1,2}
\]

(B.9)

take the derivative with respect to \( \beta_1 \)

\[
\frac{\partial q}{\partial \beta_1}(\beta_1, p_1, \bar{q}(\beta_1, p_1, p_{1,2}, p_2), p_2) + \frac{\partial q}{\partial \beta_2}(\beta_1, p_1, \bar{q}(\beta_1, p_1, p_{1,2}, p_2), p_2) \times \frac{\partial \bar{q}}{\partial \beta_1}(\beta_1, p_1, p_{1,2}, p_2) = 0
\]

(B.10)

Rearranging terms and since \( \frac{\partial q}{\partial \beta_s} > 0 \) for all \( s \) we have:

\[
\frac{\partial \bar{q}}{\partial \beta_1}(\beta_1, p_1, p_{1,2}, p_2) = -\frac{\frac{\partial q}{\partial \beta_1}(\beta_1, p_1, \bar{q}(\beta_1, p_1, p_{1,2}, p_2), p_2)}{\frac{\partial q}{\partial \beta_2}(\beta_1, p_1, \bar{q}(\beta_1, p_1, p_{1,2}, p_2), p_2)} < 0
\]

(B.11)

Then, since \( \frac{\partial q}{\partial \beta_s} > 0 \) and \( \frac{\partial \bar{q}}{\partial \beta_s} < 0 \) for all \( s \), it follows from equation (B.8) that \( \frac{\partial r}{\partial \beta_1} < 0 \).

B.2 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:

$$ I(h_{it}) = \sum_{s=1}^{S_t} \sum_{k=1}^{K_s} \frac{\left( \frac{\partial \Pr_{sk}}{\partial h_{it}} \right)^2}{\Pr^2_{sk}} - \frac{\Pr_{sk} \frac{\partial^2 \Pr_{sk}}{\partial h_{it}^2}}{\Pr^2_{sk}}. \quad (B.12) $$

where $\Pr_{sk} = \Pr(Z_{its} = k | h_{it})$. 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 appendix B.2 for definition of this estimator. This corresponds to the function $B(\cdot)$ in equation (5.13) of Section 5.1.2. This function is:

$$ \text{Bias}(h_{it}) = -\frac{1}{2[I(h_{it})]^2} \sum_{s=1}^{S_t} \sum_{k=1}^{K_s} \frac{\partial \Pr_{sk} \frac{\partial^2 \Pr_{sk}}{\partial h_{it}^2}}{\Pr^2_{sk}}. \quad (B.13) $$

### B.3 Alternative estimators of time investments

**Expected a posteriori Estimator**

With the estimates of $(\alpha, \beta)$, 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, $z_{it}$. 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.

$$ \tilde{g}_t(h_{it} \mid z_{it}; \alpha, \beta) = \frac{f_t(z_{it} \mid h_{it}; \alpha, \beta)g_t(h_{it})}{\int f(z_{it} \mid h_{it}; \alpha, \beta)g_t(h_{it})dh_{it}} \quad (B.14) $$

The posterior expected value estimator is defined as follows:

$$ h_{it}^{\text{EAP}} = E(h_{it} \mid z_{it}) = \int h_{it}\tilde{g}(h_{it} \mid z_{it}; \alpha, \beta)dh_{it} \quad (B.15) $$

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

**Maximum Likelihood Estimator**

Another estimator for each student’s time investment is the value that maximizes the indi-
vidual likelihood. That is,

$$h_{it}^{ML} \equiv \arg\max_{h_{it}} \log f(z_{it} \mid h_{it}; \alpha, \beta) \quad (B.16)$$

The bias of this estimator is $\text{Bias}(h_{it})$ in equation (B.13).

### C Monte Carlo: time investment 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:

$$h = \delta C + \eta \quad (C.1)$$

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 $\theta$, have a production technology with a single input $h$:

$$\theta = \gamma h + \nu \quad (C.2)$$

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 $\delta$ and $\gamma$.

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:

$$Z_s = k \quad \text{if and only if} \quad \alpha_{sk} \leq \beta_s h + \epsilon_s < \alpha_{sk+1} \quad (C.3)$$

for $k = 1, 2, \ldots, K_s$,

where $\alpha_{s1} = -\infty$ and $\alpha_{sk+1} = \infty$. The parameters of each ordered categorical questions are drawn from the following distributions: $\beta_s \sim \text{Uniform}(0, 2)$ and $\alpha_{sk} \sim N(0, \beta_s^2)$ with $k = 1, \ldots, K_s - 1$. The $K_s - 1$ realizations of $\alpha_{sk}$ 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 \( \varepsilon_s \). 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:

\[
Z_s = \tilde{\alpha}_s + \tilde{\beta}_s h + \varepsilon_s, \tag{C.4}
\]

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 6.1.2. Estimation procedures used are the weighted maximum likelihood (WML) estimator described section 7.1.2, 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 \( \delta \) and \( \gamma \) 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:

\[
h_s^L = \frac{Z_s - \tilde{\alpha}_s}{\tilde{\beta}_s} \tag{C.5}
\]

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 \( E = \text{WML, ML, EP} \) 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 C1 presents the results.

I estimate \( \gamma \) with 2SLS estimator and \( \delta \) with OLS estimator. Treating a categorical variable as continuous “adds” measurement error. However, the 2SLS estimator of \( \gamma \) deals with this measurement error and it converges to its true value. Similarly, the 2SLS estimator of \( \gamma \) using
Table C1: Monte Carlo Simulation

<table>
<thead>
<tr>
<th>Measurement error $e_s$ distribution</th>
<th>Linear $\delta = 0.5$</th>
<th>Non-linear $\delta = 0.5$</th>
<th>Linear $\gamma = 1$</th>
<th>Non-linear $\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>EP-OLS</td>
<td>ML-OLS</td>
<td>WML-OLS</td>
</tr>
<tr>
<td>logarithmic(0,1)</td>
<td>0.496 (0.095)</td>
<td>0.482 (0.017)</td>
<td>0.511 (0.019)</td>
<td>0.488 (0.018)</td>
</tr>
<tr>
<td>uniform(-3,3)</td>
<td>0.488 (0.105)</td>
<td>0.479 (0.016)</td>
<td>0.510 (0.019)</td>
<td>0.485 (0.018)</td>
</tr>
<tr>
<td>normal(0,1)</td>
<td>0.495 (0.045)</td>
<td>0.481 (0.016)</td>
<td>0.505 (0.019)</td>
<td>0.481 (0.020)</td>
</tr>
<tr>
<td>degenerate</td>
<td>0.493 (0.036)</td>
<td>0.481 (0.099)</td>
<td>0.510 (0.096)</td>
<td>0.484 (0.103)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Note: Monte Carlo simulation. Standard deviations in parentheses. All non-linear model are estimated under the assumption than $e_s \sim \text{logistic}(0,1)$. Estimators of each individual $h$: weighted maximum likelihood (WML), maximum likelihood (ML), expected a posteriori (EP). Estimators of $\gamma$ and $\delta$: ordinary least squares (OLS) and two-stage least squares (2SLS).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 $\delta$. 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.

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 $\Omega_{it} = \{\theta_{it}, C_{it}, x_{it}, z_{it}\}$ be the state space vector, where the state variables are child’s skill $\theta_{it}$, classroom inputs $C_{it}$, a vector of demographic characteristics $x_{it}$ (including household income $y_{it}$) and other observable inputs $z_{it}$ in the skill formation. The classroom and income processes—known by the household—are $\log C_{it} = \kappa_i' x_{it} + \Delta_{it}$ and $\log y_{it} = \bar{y}_t + \rho_t \log y_{it} + \omega_{it}$. When household make decisions they know $\Delta_{it}$ and $\omega_{it}$, but they don’t know their in subsequent grades or the current skill shock $v_{it}$—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 $\Delta_{it}$, and Gauss–Hermite quadratures for $\omega_{it}$ and $v_{it}$. 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 $\Delta_{it}$ from the assumed distribution. Second, I draw randomly $S = 200$ points of support the rest of the variables in $\Omega_{it}$. Third, I simulate the household choices, calculate the value function, and take the average across the $D$ realizations of $\Delta_{it}$ and the Gauss–Hermite quadrature integration of $\omega_{it}$ and $v_{it}$. The end result is the expected value of the value function. Lastly, I regress that expected value function on the values drawn of $\Omega_{it}$ (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.
Supplemental Tables and Figures

### Table S1:
Classroom and teacher value added

<table>
<thead>
<tr>
<th>School grade</th>
<th>Fourth</th>
<th>Sixth</th>
<th>Eighth</th>
<th>Tenth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom VA</td>
<td>0.39</td>
<td>0.26</td>
<td>0.31</td>
<td>0.26</td>
</tr>
<tr>
<td>Classrooms</td>
<td>31,639</td>
<td>31,499</td>
<td>38,102</td>
<td>38,438</td>
</tr>
<tr>
<td>Students</td>
<td>721,942</td>
<td>719,470</td>
<td>876,631</td>
<td>882,752</td>
</tr>
<tr>
<td>Teacher VA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>general</td>
<td>0.36</td>
<td>0.21</td>
<td>0.28</td>
<td>0.19</td>
</tr>
<tr>
<td>within school</td>
<td>0.18</td>
<td>0.13</td>
<td>0.22</td>
<td>0.17</td>
</tr>
<tr>
<td>Teachers</td>
<td>3,362</td>
<td>3,174</td>
<td>5,268</td>
<td>4,674</td>
</tr>
<tr>
<td>Classrooms</td>
<td>9,153</td>
<td>8,566</td>
<td>19,229</td>
<td>16,413</td>
</tr>
<tr>
<td>Schools</td>
<td>1,478</td>
<td>1,380</td>
<td>2,706</td>
<td>2,324</td>
</tr>
<tr>
<td>Students</td>
<td>238,681</td>
<td>224,218</td>
<td>462,374</td>
<td>403,760</td>
</tr>
</tbody>
</table>

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 7.2.
Figure S1:
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 7.3.
Figure S2:
Sample average marginal effects of skill formation inputs (in SD log skill 2nd grade) by mother education

Parental time (weekly hours)  
Child effort (weekly hours)  
Classrooms (SD)  
Current skills (SD 2nd grade)

Note: The values on these graphs present the average marginal effect calculated using the estimates from the specifications in Table 8. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student’s marginal effect using each input analogous equation (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 S3:
Sample average marginal effects of skill formation inputs (in SD log skill 2nd grade) by household income quintiles

Parental time (weekly hours)

Child effort (weekly hours)

Classrooms (SD)

Current skills (SD 2nd grade)

Note: The values on these graphs present the average marginal effect calculated using the estimates from the specifications in Table 8. The grey area reports school network-clustered bootstrapped 95% confidence intervals. I compute each student’s marginal effect using each input analogous equation (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 S4:
Parental time and child effort responses
to reassignment from 10th to 90th percentile of classroom and teacher distributions

Parental time (weekly hours)

Classrooms

Teachers

Child effort (weekly hours)

Classrooms

Teachers

Note: The values on the graphs are calculated using the estimates of Table 12. The top and bottom panels report 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 7.3.
Figure S5:
Household responses by child gender (F=female, M=male) to reassignment from 25th to 75th percentile of classroom distributions

Parental time (weekly hours)

Child effort (weekly hours)

Note: The values on the graphs are calculated using the estimates of Table 12. 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 7.3.
Figure S6:
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 9.2 for additional details.
Figure S7:
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 8). 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 (8.16). 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 9.2 for additional details.