PERUVIAN ECONOMIC ASSOCIATION

Too Hard, Too Easy, or Just Right: Dynamic Complementarity and Substitutability in The Production of Skill

Juan F. Castro

Lucciano Villacorta

Working Paper No. 175, December 2020

Too Hard, Too Easy, or Just Right: Dynamic Complementarity and Substitutability in The Production of Skill

Juan F. Castro and Lucciano Villacorta[1]

Evidence shows that skill attained early in the life of children can either increase (dynamic complementarity) or reduce (dynamic substitutability) the effect of inputs occurring later. We propose a novel production function of cognitive skill that allows the same input to exhibit both phenomena using the notion that learning is maximized when the skill already attained by the child matches the complexity of the input. We estimate this function using panel data on test scores and schooling for a large sample of Peruvian children, and a non-linear version of the Arellano-Bond GMM estimator. We find schooling exhibits dynamic complementarity when its complexity exceeds the child's skill and dynamic substitutability otherwise.

JEL Codes: O15, C33

Keywords: production function of skill, dynamic complementarity, dynamic substitutability.

---

[1] Juan F. Castro is a professor of Economics at Universidad del Pacifico (e-mail: castro_jf@up.edu.pe). Lucciano Villacorta is Senior Economist at the Economic Research Department of the Central Bank of Chile (e-mail: lvillacorta@bcentral.cl). Authors would like to thank Manuel Arellano, Stéphane Bonhomme, Alonso Villacorta and participants at the Economics Research Seminar at Universidad del Pacifico, the 2019 Annual Meeting of the Peruvian Economic Association and the 2019 Latin American Meeting of the Econometric Society for their comments. We also thank Alexandra Heredia for excellent research assistance. All remaining errors are ours. The data used in this article is available online in the UK Data Archive at https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000060. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

1. Introduction

Skills attained during childhood play a key role in the process of human capital formation. Numerous studies have found a strong causal relationship between the skills developed during childhood and later-life outcomes such as post-secondary schooling, employment status, wages, and participation in crime (Almond and Currie (2011); Cunha *et al.* (2006)). Economics has contributed to the understanding of the skill formation process by formulating production functions and estimating the effect of inputs occurring in different places (e.g. homes and schools) and at different stages (e.g. during early childhood and at school-age).

A prominent strand of this literature has found evidence of dynamic complementarity in the production of cognitive skill (Cunha and Heckman (2008); Cunha et al. (2010); Aizer and Cunha (2012)). This means that skill attained at early stages in the life of children increases the effect of inputs occurring later. This finding is one of the key arguments put forward to highlight the importance of early life investments to achieve a more capable and productive adult population (Heckman and Mosso (2014)).

In a very influential study within this literature, Cunha et al. (2010) estimated the effect of parental investments (material goods and time spent with their children) on cognitive skill using data from the National Longitudinal Survey of the Youth (NLSY79) in the United States. They found that early life investments are not only more productive than those occurring later but also have a positive effect on the productivity of future investments.

The logic behind dynamic complementarity is that children that have been exposed to a more nurturing environment during their early years are better prepared for the learning experiences they encounter later. More recent empirical work, however, has found evidence that an increase in children's previous cognitive attainment can reduce the productivity of an input, and has labeled this as *dynamic substitutability*. Agostinelli and Wiswall (2016) and Garcia and Gallegos (2017) estimated the productivity of parental investments using data from the NLSY79 and the sample of children participating in the Infant Health and Development Program[2], respectively. Contrary to the findings of Cunha *et al.* (2010), they both found that a raise in children's prior stock of skill has a detrimental effect on the productivity of parental investments.

In this paper, we estimate a flexible production function of cognitive skill that reconciles the evidence surveyed above by allowing the same input to exhibit dynamic complementary and substitutability. We explain this heterogeneity using the notion that learning is maximized when there is a match between two variables. One is the cognitive demand or complexity of the input. The other is the level of cognitive skill that the child brings to the interaction with the input. Deviations from this match in either direction are detrimental for the productivity of the input.

Whether an input exhibits dynamic complementarity or substitutability will be determined by the difference between the cognitive demand of the input and the cognitive skill of the child. A positive difference favoring the cognitive demand of the input will lead to dynamic complementarity. This is because an increase in the cognitive skill of the child will reduce the mismatch and, thus, raise the productivity of the input. A negative difference will produce dynamic substitutability because, in this case, raising the skill of the child will widen the mismatch.

[2] The Infant Health and Development Program was a large-scale randomized-controlled trial that began in 1985 and took place in eight states of the United States.

For a given input, one can think of the mismatch between the cognitive skill of children and the complexity of the input as a misallocation of child skill. This can occur for several reasons. For example, incomplete information about child skill or financial frictions can prevent parents from purchasing the most appropriate input. Also, if one thinks of the interactions that occur at school as an input of skill (as we do in this paper), it is not hard to imagine that mismatches will occur because these interactions cannot be perfectly tailored to every student.

The idea that learning is enhanced when the experience or stimulus is appropriate for the learner's degree of understanding is present across different theories of learning (see, for example, Twomey and Stewart (2005) on Constructivism and Paas *et al.* (2010) on Cognitive Load Theory). Recent research on the economics of human development also acknowledges the importance of offering the child experiences that are neither too hard or too easy to avoid discouraging him/her (see, for example, Heckman and Mosso (2014) on the strategy of "scaffolding"). On the empirical front, there is evidence showing that the pedagogical approach known as "Teaching at the Right Level" ─which requires adapting the interactions proposed in class to the competency level of students─ can produce positive effects on learning (see Banerjee *et al*. (2016)).

Despite the theoretical considerations and the empirical evidence cited above, the notion that learning is enhanced when there is a match between the child's ability and the complexity of the input has not yet been introduced into the technology of skill formation. This is the first study to propose a production function of cognitive skills that incorporates this idea, to test it using data, and to use it to reconcile the evidence regarding the coexistence of dynamic complementarity and substitutability.

In our empirical production function, cognitive skill is influenced by the number of years of schooling, school characteristics, and the child's prior stock of cognitive skill. A key aspect of our modeling approach is to allow for individual-specific effects of schooling that depend on the distance between the child's prior skill attainment and the cognitive demand of the school (the mismatch).

To identify the parameters that govern the production function of skills, we rely on observed heterogeneity in these mismatches. We include individual fixed effects in the production function to control for potential unobserved heterogeneity that might be correlated with schooling and the stock of skills, and take advantage of three rounds of data to estimate a non-linear dynamic panel model. For this, we introduce a non-linear version of the Arellano-Bond GMM estimator that exploits valid moment conditions.

We use longitudinal information on test scores, years of schooling and school characteristics for a large sample of Peruvian children attending to different schools. We measure cognitive skill using the scores obtained by the children in the Peabody Picture Vocabulary Test. Differences in cognitive demand across schools are approximated using the heterogeneity in the proportion of the Mathematics curriculum that is covered in each school as well as the average performance of the child's school peers in a Mathematics test.

We find empirical evidence that supports the claim that the same input can exhibit both dynamic complementarity and substitutability depending on the difference between the child's past skill attainment and the cognitive demand of the input. In fact, we find that in the school with the median level of cognitive demand, the relation between the effect of schooling and the child's prior stock of skill follows an inverted-U shape. The input exhibits dynamic complementarity for almost the entire first half of the distribution of prior skill and dynamic substitutability for its second half.

Consistent with our framework, dynamic substitutability dominates if we reduce the complexity of the input. In fact, we observe dynamic substitutability for most of the support of prior skill in the school at the 25^{th} percentile of cognitive demand. The contrary happens if one increases the complexity of the input. In particular, we observe dynamic complementarity for most of the support of prior skill in the school at the 75^{th} percentile of cognitive demand. Importantly, all these results are robust to the two different measures of cognitive demand considered.

The rest of the paper is organized as follows. Section 2 presents the production function proposed to accommodate the notion that input productivity depends on the child's prior skill but relative to the complexity or cognitive demand of the input. Section 3 discusses the data and empirical strategy employed to identify the productivity of schooling. We present our results in Section 4. Section 5 closes with some concluding remarks.

2. Framework

We propose a production function of skill that incorporates the notion that the productivity of an input is maximized when there is match between the cognitive demand or complexity of the input and the cognitive skill that the child brings to the interaction with the input. Deviations from this match mean that the input is either too complex or too simple for the child and, thus, deviations are detrimental for its effect on learning (i.e. deviations are detrimental for the productivity of the input).

Let us assume the following production function of skill:

$$A_{it} = A_{it-1}^{\gamma_1} I_{it}^{\phi_{it}} e^{\alpha_i + \mu_{it}} \tag{1}$$

This has the following log-linear version:

$$a_{it} = \gamma_1 a_{it-1} + \phi_{it} i_{it} + \alpha_i + \mu_{it} \tag{2}$$

where a_{it} represents some measure of cognitive skill of child i at time t, i_{it} indicates exposure to the input, α_i captures unobserved heterogeneity that is allowed to be correlated with all the observables of the model, and μ_{it} represents all other unobservable variables that are not correlated with the input or the child's prior skill attainment. The parameter γ_1 captures persistence in cognitive skills.

The productivity of the input is given by ϕ_{it}. This is, the effect on skill of an additional unit of exposure to the input. Note that this productivity is individual-specific and, in particular, will be allowed to vary depending on the difference between the level of cognitive skill previously attained by the child (a_{it-1}) and the level of cognitive demand of the input to which the child is exposed (D_i).

Consider the following functional form for ϕ_{it}:

$$\phi_{it} = \frac{\gamma_2}{\exp[\lambda(a_{it-1} - \theta D_i)^2]} \tag{3}$$

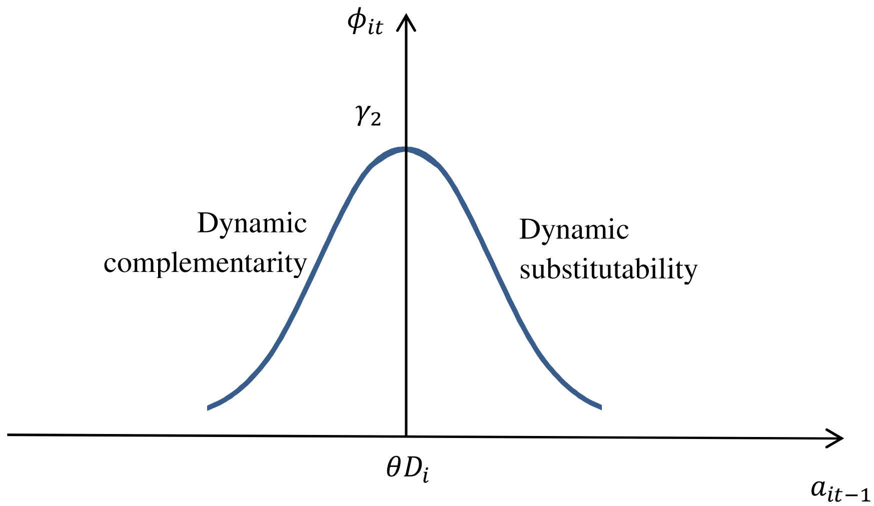
In (3), the productivity of the input is maximized at a value of γ_2 when there is a match between the cognitive demand of the input and the child's prior cognitive skill $(a_{it-1} = \theta D_i)$. Differences between a_{it-1} and θD_i introduce variations in ϕ_{it}. In fact, the larger these differences, the smaller the productivity. Parameter θ transforms units of cognitive demand of the input D_i in units of stock of skills a_{it-1}. Importantly, it also allows our production function to nest models that only exhibit either dynamic complementarity or dynamic substitutability (see discussion below). Parameter λ controls the curvature of ϕ_{it}. In the extreme case where $\lambda = 0$, the productivity of the input is constant and equal to γ_2.

Notice that the effect of the child's prior skill attainment on the productivity of the input is given by:

$$\frac{\partial \phi_{it}}{\partial a_{it-1}} = \frac{-\gamma_2 2\lambda(a_{it-1} - \theta D_i)}{\exp[\lambda(a_{it-1} - \theta D_i)^2]} = -\phi_{it} 2\lambda(a_{it-1} - \theta D_i) \tag{4}$$

Therefore, if both parameters γ_2 and λ have a positive sign, the input will exhibit dynamic complementarity $\left(\frac{\partial \phi_{it}}{\partial a_{it-1}} > 0\right)$ when $a_{it-1} - \theta D_i < 0$. Intuitively, the input is too complex for the child so raising his/her skill will enhance the effect of the input on learning. In the opposite case, when $a_{it-1} - \theta D_i > 0$, the input will exhibit dynamic substitutability $\left(\frac{\partial \phi_{it}}{\partial a_{it-1}} < 0\right)$. Intuitively, the input is too easy for the child so raising his/her skill will be detrimental for the input's productivity. Figure 1 illustrates this.

Figure 1
Dynamic complementarity and substitutability
in the same input of cognitive skill



It is important to notice that the framework in (3) nests the cases where the production function only exhibits either dynamic complementarity or dynamic substitutability. For example, if $\theta = 0$ and $\lambda < 0$, the production function always displays dynamic complementarity, since $\frac{\partial \phi_{it}}{\partial a_{it-1}}$ is always positive. In contrast, if $\theta = 0$ and $\lambda > 0$, the production function always displays dynamic substitutability, since $\frac{\partial \phi_{it}}{\partial a_{it-1}}$ is always negative.

The idea that the productivity of an input of skill is maximized when there is match between its complexity and the cognitive skill of the learner is consistent with several theories of learning. For example, Constructionism proposes that learning occurs within the child's "zone of proximal development", which reflects the ability of the child to understand the logic of the new concept (Twomey and Stewart (2005)).

Cognitive Load Theory (CLT) is another influential theory of learning that underscores the importance of achieving a balance between the complexity of the stimulus and the expertise of the learner. CLT is based on the notions that the human working memory has a limited capacity to process new information and that learning requires processing, organizing and storing this new information in long-term memory (Leppink *et al.* (2015)). Two important ideas behind this theory are: (i) that the amount of cognitive load produced by an educational process depends on the difference between the complexity of the information being presented and the learner's prior knowledge; and (ii) that learning requires a certain amount of cognitive load (too little or too much cognitive load are both detrimental for learning).

Combining these two ideas one deduces that, for a given degree of complexity of the information to be processed, learning is enhanced at a certain level of skill (an excess above this level or a deficit below it, are both detrimental for learning). Moreover, CLT has two implications for our setting. First, that there must be a negative relation between the amount of cognitive load experienced by child i and the difference $a_{it-1} - \theta D_i$. Second, that the amount of cognitive load produced at the match $a_{it-1} = \theta D_i$ must be positive. [3]

[3] We use the available data to approximate cognitive load and test these two implications. The results are presented in the Appendix.

3. Data and empirical strategy

3.1 Data

We use the information of the Younger Cohort of the Young Lives Study in Peru.[4] In particular, we use rounds 2, 3 and 4 of the child survey collected between 2006 and 2013, and the school survey collected in 2011. The basic structure of this data is summarized in Table 1. We use the sample of children that have complete cognitive test scores for rounds 2, 3 and 4, and attend a school that participated in the school survey (480 children in 125 schools).[5]

Table 1
Structure and sample sizes of the relevant Young Lives databases

	Child survey			**School Survey 2011**
	Round 2 2006	**Round 3 2009**	**Round 4 2013**	
Younger cohort's age (years)	5	8	12	10
Sample size (children)	2,052	1,943	1,902	572 (132 schools)
Educational attainment	Preschool	Grade 2	Grade 6	Grade 4

Source: Young Lives Study (Peru).

Following (2) and (3), we need information on three variables to estimate the parameters of the production function: the cognitive skill of the child, his/her exposure to an input, and the cognitive demand of this input. We will use the test scores obtained in the Peabody Picture Vocabulary Test as a measure of the cognitive skill of the child.

[4] Young Lives is an international study of childhood poverty, following 12,000 children in 4 countries (Ethiopia, India, Peru and Vietnam) over 15 years.

[5] The risk of selection bias due to this second condition is very small. Primary school attendance in Peru is close to 100% (only 0.7% and 0.3% of Young Lives younger cohort children were not attending school in round 3 and round 4, respectively) and schools participating in the school survey were randomly selected (Guerrero *et al.* (2012)).

This is a widely used test of receptive vocabulary that has a strong positive correlation with several measures of intelligence (Cueto and Leon (2012)). The input considered in this analysis are the influences originated at school. Therefore, we will use the number of school years attained by the child to reflect the degree of exposure to the input.

Similar to the skill of the child, the cognitive demand of an input is a latent variable. Intuitively, one can relate the cognitive demand of a school input to the depth and quantity of topics covered during classes. We will use the school survey collected in 2011 to provide a proxy for the average cognitive demand of the school inputs to which the children were exposed between rounds 2 and 4.

We will consider two different variables to reflect this average cognitive demand. The first is the heterogeneity in the Mathematics curriculum coverage reported in the class attended by the child.[6] The complexity of school curricula usually increases as more topics are covered and this is particularly true for Mathematics, where topics are sequenced ranging from simple to more complex. The second variable is the average performance of the child's school peers in a Mathematics test administered as part of the Young Lives School Survey.[7] The logic is that teachers manage the complexity of the interactions and information presented during class considering the proficiency of the

[6] For the School Survey, the Mathematics teachers of the Young Lives children were given a comprehensive list of topics and asked how many of them have been covered in class. The curriculum coverage corresponds to the proportion of these topics reported as covered in depth. We use the average coverage if there is more than one Young Lives child in the same school.

[7] For the School Survey, a random sample of class peers was chosen for each Young Lives child. A maximum of two classes and 20 peers were randomly selected per school, depending on the class size and the number of classes in which the Young Lives children were enrolled within the same school. These class peers took a Mathematics test that measured their numeracy skills.

group. This behavior has already been documented in the literature (see, for example, Duflo *et al.* (2011)).

Table 2 presents descriptive statistics for all the variables used in the empirical specification.

Table 2
Descriptive statistics of the variables used in the empirical specification

Variable used in the empirical specification	Round	Mean	SD
Raw PPVT score	2	29.22	16.94
	3	60.60	16.67
	4	86.30	16.48
Years of schooling	3	2.37	0.54
	4	5.00	0.05
Proportion of the Mathematics curriculum covered	School Survey	0.488	0.16
Average school peers' Mathematics standardized test scores	School Survey	-0.03	0.69

Note: the number of observations is 480 for all variables.

3.2 Empirical strategy

The model we bring to the data is the log of the production function described in equation (2). Equation (2) is a linear model but with a heterogeneous effect and can be framed within the potential outcome framework with heterogeneous effects of the "treatment" i_{it}. One difference in our specification is that we are parametrizing the individual-specific effect as a function of the mismatch between the stock of skill of child i and the cognitive demand of the treatment. Therefore, we are not only interested in identifying a local average treatment effect but how the marginal effect of schooling evolves along the entire distribution of skills.

Following Andrabi *et al.* (2011), we take advantage of the fact that our database possesses three waves of data to include an individual fixed-effect in our empirical model. The fixed-effects control for any unobservable characteristic that might influence skills and be correlated with the treatment (i_{it}), the stock of skill when the treatment takes place (a_{it-1}), or the cognitive demand of the treatment (D_i).

In general, we can expect a strong correlation between an observed input of skill and other unobserved influences both at the family and school level. More affluent families are not only capable of purchasing more and better school inputs but they are also capable of offering a more nurturing environment at home and during early childhood. The fixed-effects identification in our panel data model requires that these unobservable characteristics do not change between the last two waves of data. This is a reasonable assumption insofar unobserved determinants of skill are likely related to family resources and preferences and the last two waves cover a relatively short period of time (4 years).

Our empirical model is as follows:

$$a_{it} = \gamma_1 a_{it-1} + \frac{\gamma_2}{\exp\left[\lambda\left(\tilde{a}_{it-1} - \theta\widetilde{D}_i\right)^2\right]} i_{it} + \alpha_i + \mu_{it} \tag{5}$$

Where $\tilde{a}_{it-1}$is the standardized value of a_{it-1} and $\widetilde{D}_i$ is the standardized proportion of the Mathematics curriculum covered in the class attended by child i when the school survey was collected or the standardized average score attained by the school peers of child i in the Mathematics test administered during the school survey.[8]

[8] Past cognitive skill and the two proxies of cognitive demand are each measured in a different scale. We use their standardized values to make them comparable.

Given the empirical specification in (5), we need to estimate four parameters $\gamma_1, \gamma_2, \lambda, \theta$. Notice that the model in (5) is a non-linear dynamic panel model. The inclusion of the fixed-effects and the lagged value of the stock of skill generates an incidental parameter problem for fixed-T if we estimate equation (5) through a non-linear version of the within group estimator (or a non-linear version of the model in first difference for $T = 2$). To see this, consider the equations for $t = 3$ and $t = 2$:

$$a_{i3} = \gamma_1 a_{i2} + \phi_{i3} i_{i3} + \alpha_i + \mu_{i3} \tag{6}$$

$$a_{i2} = \gamma_1 a_{i1} + \phi_{i2} i_{i2} + \alpha_i + \mu_{i2} \tag{7}$$

We can take the first difference to eliminate the unobserved heterogeneity α_i:

$$a_{i3} - a_{i2} = \gamma_1 (a_{i2} - a_{i1}) + \phi_{i3} i_{i3} - \phi_{i2} i_{i2} + \mu_{i3} - \mu_{i2} \tag{8}$$

Notice that:

$$\frac{\partial \phi_{it}}{\partial \gamma_2} = \frac{1}{\exp\left[\lambda\left(\tilde{a}_{it-1} - \theta \widetilde{D}_i\right)^2\right]} \tag{9}$$

$$\frac{\partial \phi_{it}}{\partial \lambda} = \frac{-\gamma_2\left(\tilde{a}_{it-1} - \theta \widetilde{D}_i\right)^2}{\exp\left[\lambda\left(\tilde{a}_{it-1} - \theta \widetilde{D}_i\right)^2\right]} \tag{10}$$

$$\frac{\partial \phi_{it}}{\partial \theta} = \frac{\gamma_2 2\lambda \widetilde{D}_i\left(\tilde{a}_{it-1} - \theta \widetilde{D}_i\right)}{\exp\left[\lambda\left(\tilde{a}_{it-1} - \theta \widetilde{D}_i\right)^2\right]} \tag{11}$$

Thus, the non-linear OLS estimator uses the sample analogue of the following optimal moment conditions:

$$E\begin{bmatrix} (\mu_{i3}-\mu_{i2})(a_{i2}-a_{i1}) \\ (\mu_{i3}-\mu_{i2})\left(\frac{i_{i3}}{\exp\left[\lambda\left(\tilde{a}_{i2}-\theta\widetilde{D}_i\right)^2\right]}-\frac{i_{i2}}{\exp\left[\lambda\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2\right]}\right) \\ (\mu_{i3}-\mu_{i2})\left(\frac{-\gamma_2\left(\tilde{a}_{i2}-\theta\widetilde{D}_i\right)^2 i_{i3}}{\exp\left[\lambda\left(\tilde{a}_{i2}-\theta\widetilde{D}_i\right)^2\right]}-\frac{-\gamma_2\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2 i_{i2}}{\exp\left[\lambda\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2\right]}\right) \\ (\mu_{i3}-\mu_{i2})\left(\frac{\gamma_2 2\lambda\widetilde{D}_i\left(\tilde{a}_{i2}-\theta\widetilde{D}_i\right)i_{i3}}{\exp\left[\lambda\left(\tilde{a}_{i2}-\theta\widetilde{D}_i\right)^2\right]}-\frac{\gamma_2 2\lambda\widetilde{D}_i\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)i_{i2}}{\exp\left[\lambda\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2\right]}\right) \end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix} \tag{12}$$

The moment conditions in (12) do not hold because of correlation between a_{i2} and μ_{i2}. This is a non-linear version of the well-known Nikell bias that appears in the dynamic linear panel data model for fixed T. To overcome this problem, we take advantage of a third round of data and we claim $T = 3$ identification by using a non-linear version of the Arellano-Bond GMM estimator. The moment conditions we employ, therefore, are as follows:

$$E\begin{bmatrix} (\mu_{i3}-\mu_{i2})(-a_{i1}) \\ (\mu_{i3}-\mu_{i2})\left(-\frac{i_{i2}}{\exp\left[\lambda\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2\right]}\right) \\ (\mu_{i3}-\mu_{i2})\left(-\frac{-\gamma_2\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2 i_{i2}}{\exp\left[\lambda\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2\right]}\right) \\ (\mu_{i3}-\mu_{i2})\left(-\frac{\gamma_2 2\lambda\widetilde{D}_i\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)i_{i2}}{\exp\left[\lambda\left(\tilde{a}_{i1}-\theta\widetilde{D}_i\right)^2\right]}\right) \end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix} \tag{13}$$

4. Results

Table 3 presents the estimates of the four production function parameters involved in (5), considering the two measures of cognitive demand.

Table 3
Production function parameters

	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\lambda}$	$\hat{\theta}$
Mathematics curriculum covered	0.2922*** (0.0448)	0.5875*** (0.1660)	0.4143*** (0.1040)	1.5604*** (0.2966)
School peers' Mathematics test scores	0.2971*** (0.0435)	0.4532*** (0.1312)	0.3469*** (0.0940)	1.2268*** (0.4406)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

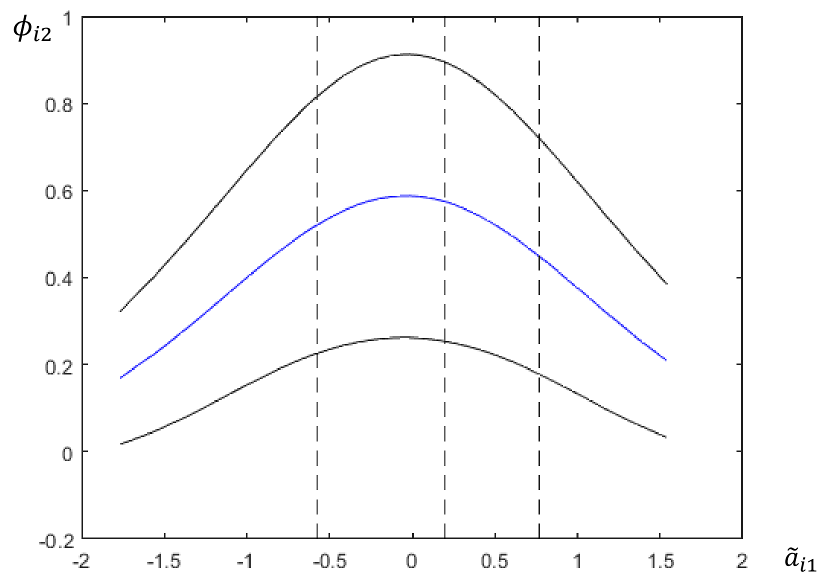
Notice that all estimates are statistically significant and point estimates remain robust to the variable used to reflect the cognitive demand of the school input D_i. Parameter γ_1 captures the persistence of skill. Our results show that around 30% of skill is carried forward from one round to the next. The estimated persistence parameter is in line with the results found in Andrabi *et al.* (2011), who use dynamic panel data techniques to estimate a linear version of the skill production function with unobserved heterogeneity. The Arellano-Bond estimates of the persistence parameter found by Andrabi *et al.* (2011) lie between 0.12 and 0.35 for different test scores.

Recall that γ_2 is the maximum productivity of schooling. This is, the productivity attained when there is match between the cognitive skill of the child and the cognitive demand of the input. According to the results presented in Table 3, a 1% increase in schooling can produce up to a 0.45-0.59% increase in cognitive skill. Departures from the match will reduce this marginal effect. It is also worth noticing from Table 3 that the estimated value of λ is statistically different from zero. This means there is heterogeneity in the effect of schooling as $\lambda = 0$ would imply a constant productivity given by γ_2. In addition, notice that parameter θ is also different from zero. This means the data does not support a production function that only exhibits either dynamic complementarity or dynamic substitutability. In fact, because both parameters γ_2 and λ have a positive sign, the input will exhibit dynamic complementarity when $\tilde{a}_{it-1} - \theta\widetilde{D}_i < 0$ and dynamic substitutability when $\tilde{a}_{it-1} - \theta\widetilde{D}_i > 0$.

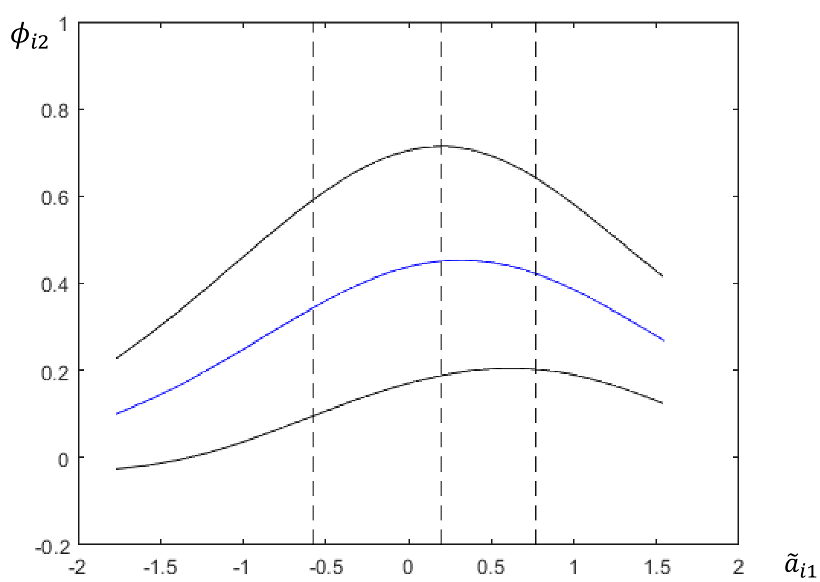
In addition to these results, we would like to evaluate how the productivity of a particular input changes with the prior skill of the child. For this, we will fix the cognitive demand of the school and use the functional form given in (5) and the parameter estimates given in Table 3 to evaluate how the productivity of an additional year of schooling changes for different values of period 1 skill ($\tilde{a}_{i1}$). The results obtained for the input with the median cognitive demand (fixing $\hat{\theta}\widetilde{D}_i = \hat{\theta}\widetilde{D}_{50}$) are presented in Figure 2. The figure shows point estimates and a 95% confidence interval. Vertical lines indicate the cut-off values of the quartiles in the distribution of $\tilde{a}_{i1}$.

Figure 2
Relation between the productivity of the school with the median cognitive demand and period 1 cognitive skill

(A) Mathematics curriculum covered



(B) School peers' Mathematics test scores



Notes: Figures include 95% confidence intervals. Standard errors computed using the delta method. Vertical lines indicate the cut-off values of the quartiles in the distribution of $\tilde{a}_{i1}$.

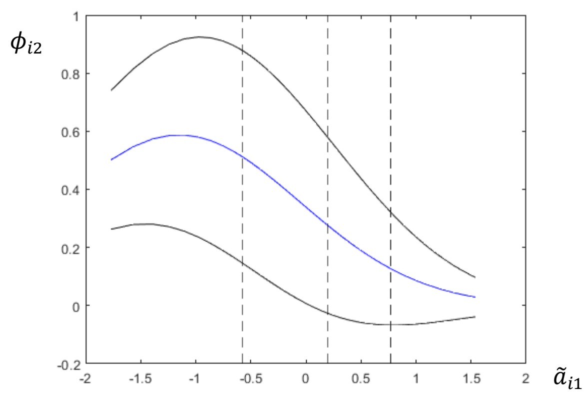
Figure 2 reveals that the productivity of schooling in a school with the median cognitive demand is a non-monotonic function of the child's past cognitive skill. The input exhibits dynamic complementarity for most of the first half of the distribution of period 1 skill. In fact, for all the children situated in the lower quartile of period 1 skill, a raise in their ability would enhance the productivity of the median school. According to the framework explained in Section 2, this raise in past skill has a positive effect on productivity because it reduces the gap between the cognitive skill of the child and the cognitive demand of the interactions proposed in the median school. A child in the lower quartile of the skill distribution would experience a larger expansion in his/her cognitive skill from interacting with this input if he/she brings more cognitive skill to this interaction.

For a school with the median cognitive demand, however, there is a limit to this positive effect of past skill on the productivity of schooling. In fact, productivity reaches the maximum of 0.45-0.59% at the 40th-50th percentile of past skill. At this point, the gap between the cognitive skill of the child and the cognitive demand of the input is zero. For higher values of past skill and most of the upper half of its distribution, the school input exhibits dynamic substitutability. This means that raising children's past skill will reduce the productivity of schooling. This is because a raise in the child's past skill increases the gap between his/her cognitive skill and the cognitive demand of the input.

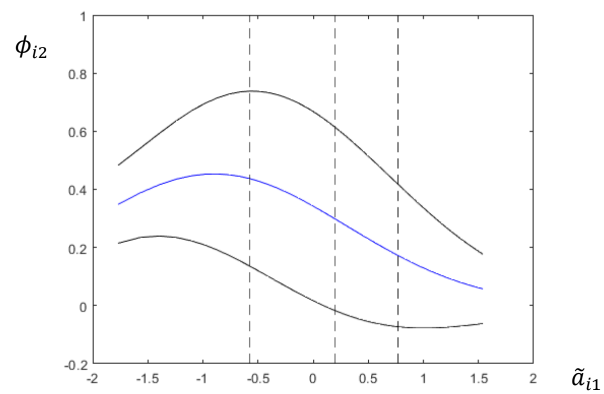
In what follows, we replicate this exercise for a school located in the 25th and 75th percentile of the cognitive demand distribution. The results are presented in Figure 3. As expected, dynamic substitutability dominates for the school in the 25th percentile of the cognitive demand distribution and dynamic complementarity dominates for the school in 75th percentile. Relative to the skill attained by the majority of children by the time they start school (period 1), the interactions proposed in the first school are too simple. Raising children's period 1 skill would increase the mismatch and reduce the effect of an additional year of schooling. The contrary is observed in a school located in the 75th percentile of cognitive demand. Interactions are more complex so, for the majority of children, raising their period 1 skill will close the mismatch and enhance the effect of an additional year of schooling.

Figure 3
Relation between the productivity of schooling and period 1 cognitive skill

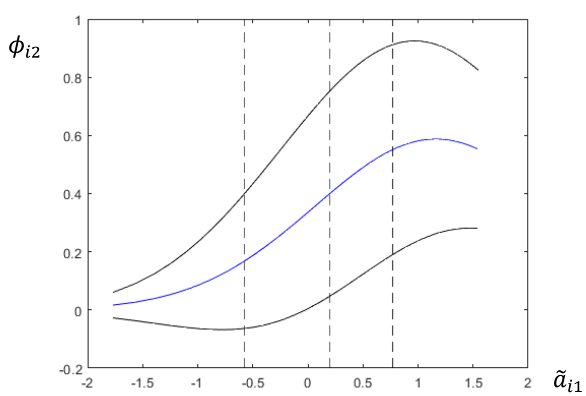
(A) Mathematics curriculum covered and 25th percentile of cognitive demand



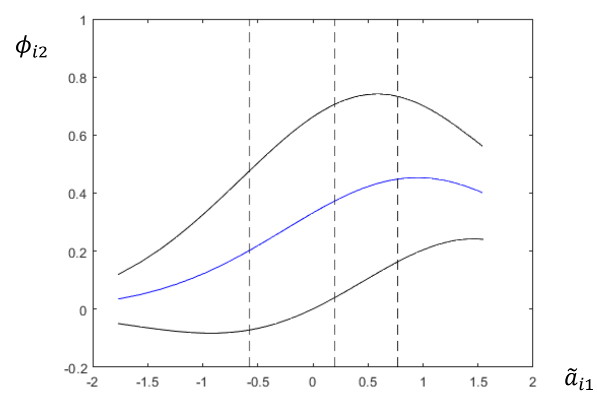
(B) School peers' Mathematics test scores and 25th percentile of cognitive demand



(C) Mathematics curriculum covered and 75th percentile of cognitive demand



(D) School peers' Mathematics test scores and 75th percentile of cognitive demand



Notes: Figures include 95% confidence intervals. Standard errors computed using the delta method.

5. Concluding remarks

We proposed a flexible production function of skill that allows the same input to exhibit both dynamic complementarity and substitutability. For this, we employed the notion that learning (input productivity) is maximized when the level of cognitive skill already attained by the child matches the complexity or cognitive demand of the input. We tested this function using longitudinal information of cognitive test scores and schooling attained by a large sample of Peruvian children. The empirical evidence found shows that the same input can exhibit dynamic complementarity when its complexity exceeds the skill of the child, and dynamic substitutability when its complexity is below the child's skill.

These findings serve to reconcile the results found in the literature, which show evidence of both dynamic complementarity and substitutability in the production of skill among children. Some authors have argued that dynamic complementarity and substitutability are characteristics of a particular input and not a general trait of the production function of skill (Garcia and Gallegos (2017)). This allows different inputs to have different relations with past skill, but does not explain why is this possible and does not clarify why the same input (e.g. material goods and time invested by parents) can show both dynamic complementarity (as in Cunha *et al.* (2010)) and substitutability (as in Angostinelli and Wiswall (2016) and Garcia and Gallegos (2017)).

Our framework explains dynamic complementarity and substitutability and allows the same input to exhibit both phenomena through a production function of skill where complementarity and substitutability emerge as the consequence of a mismatch between the skill of the child and the complexity of the learning experience. Rather than an inherent characteristic of how skill is produced or of a particular input of skill, our framework implies that dynamic complementarity and substitutability are phenomena that reflect a less than optimal assignment of a particular input.

If we relate the quality of an educational input to the amount of learning it produces among the children it is meant to serve, dynamic complementarity and substitutability also reflect a less than optimal quality. In fact, if an input exhibits dynamic complementarity or substitutability for a significant part of the distribution of skill of the children it is intended to serve, our framework predicts that its quality could be improved by reducing or enhancing its complexity, respectively.

References

Agostinelli, F., & Wiswall, M. (2016). Estimating the Technology of Children's Skill Formation. *NBER Working Paper No. 22442*.

Aizer, A., & Cunha, F. (2012). The Production of Human Capital : Endowments, Investments and Fertility. *NBER Working Paper No. 18429*.

Almond, D., & Currie, J. (2011). Human Capital Development before Age Five. In O. Ashenfelter, & D. Card, *Handbook of Labor Economics* (Vol. 4B, pp. 1315-1486). Elsevier.

Andrabi, T., Das, J., Khwaja, A., & Zajone, T. (2011). Do Value-Added Estimates Add Value? Accounting for Learning Dynamics. *American Economic Journal: Applied Economics, 3*(3), 29-54.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., . . . Walton, M. (2016). Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level in India. *NBER Working Paper No. 22746*.

Cueto, S., & Leon, J. (2012). Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives. *Young Lives Technical Note 25*.

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

Cunha, F., Heckman, J., & Lochner, L. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek, & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 1, pp. 697-812). Elseiver.

Cunha, F., Heckman, J., & Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica, 78*(3), 883-931.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review, 101*, 1739–1774.

García, J., & Gallegos, S. (2017). Complementarity and Substitutability in the Production of Early Human Capital.

Guerrero, G., Leon, J., Freire, S., Cueto, S., Rosales, E., Zapata, M., & Saldarriaga, V. (2012). Young Lives School Survey in Peru: Design and Initial Findings. *Young Lives Working Paper No. 92*.

Heckman, J., & Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics, 6*, 689-733.

Leppink, J., van Gog, T., Paas, F., & Sweller, J. (2015). Cognitive load theory: researching and planning teaching to maximise learning. In J. Cleland, & S. Durning (Eds.), *Researching Medical Education* (pp. 207-218).

Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educational Psychology Review, 22*(2), 115-121.

Twomey, C., & Stewart, R. (2005). Constructivism: A Psychological Theory of Learning. In Twomey, *Constructivism: Theory, Perspectives, and Practice* (pp. 8-38). New York and London: Teachers' College Press.

Appendix

In this Appendix, we explore the relation between the difference $\tilde{a}_{i1} - \hat{\theta}\tilde{D}_i$ and cognitive load. Cognitive load theory postulates that learning requires some cognitive load and that this load depends on the difference between the complexity of the new information and the learner's prior knowledge. This has two implications for our setting. First, that there must be a negative relation between the amount of cognitive load experienced by child i and the difference $\tilde{a}_{i1} - \hat{\theta}\tilde{D}_i$. Second, that the amount of cognitive load produced at the match $\tilde{a}_{i1} = \hat{\theta}\tilde{D}_i$ (where learning is maximized) must be positive.

We approximate cognitive load using a six-item questionnaire that captures the child's perception about the degree of difficulty of Mathematics classes, collected as part of the Young Lives School Survey. [9] We regress this measure of cognitive load on the difference $\tilde{a}_{i1} - \hat{\theta}\tilde{D}_i$ for both measures of $\tilde{D}_i$ (the proportion of the Mathematics curriculum covered and the school peers' Mathematics test scores). Results reported below are consistent with the predictions of cognitive load theory: we find a positive and significant intercept (which means that at the match $\tilde{a}_{i1} = \hat{\theta}\tilde{D}_i$ the amount of cognitive load is positive) and a negative slope relating cognitive load with $\tilde{a}_{it-1} - \hat{\theta}\tilde{D}_i$ (albeit it is only statistically significant for one of the measures of $\tilde{D}_i$).

[9] The child had to express how often (almost never, sometimes or almost always): (i) work in Mathematics classes is easy for him/her; (ii) he/she learns things quickly in Mathematics classes; (iii) he/she looks forward to Mathematics classes; (iv) Mathematics classes are interesting for him/her; (v) he/she likes Mathematics class; and (vi) he/she enjoys doing work in Mathematics classes.

These results support the use of the proportion of the Mathematics curriculum covered and the school peers' Mathematics test scores to reflect the cognitive demand of the school input. They are also consistent with the negative relation found between the productivity of schooling and the difference between the child's prior skill and this cognitive demand.

Table 2.1 Relation between perceived cognitive load and the difference between the child's prior skill and the cognitive demand of his/her school $\left(\tilde{\boldsymbol{a}}_{\boldsymbol{it-1}} - \widehat{\boldsymbol{\theta}}\widetilde{\boldsymbol{D}}_{\boldsymbol{i}}\right)$

	Measure of cognitive demand $\left(\widetilde{\boldsymbol{D}}_{\boldsymbol{i}}\right)$	
	Mathematics curriculum covered (1)	School peers' Mathematics test scores (2)
$\tilde{a}_{it-1} - \hat{\theta}\widetilde{D}_i$	-0.0627*	-0.0127
	(0.0363)	(0.0581)
Constant	10.43***	10.27***
	(0.126)	(0.183)
N	471	471

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1