

Understanding the Trends in Learning Outcomes in Argentina, 2000 to 2012

Rafael de Hoyos

Peter A. Holland

Sara Troiano



WORLD BANK GROUP

Education Global Practice Group

December 2015

Abstract

This paper seeks to understand what drove the trends in learning outcomes in Argentina between 2000 and 2012, using data from four rounds of the Program for International Student Assessment. A year-specific education production function is estimated and its results used to decompose the changes in learning outcomes into changes in inputs, parameters, and residuals via microsimulations. Estimates of the production function show the importance of socioeconomic status, gender, school autonomy, and teacher qualifications to determine learning outcomes. Despite an important increase in the level of resources

invested in public education, learning outcomes in public schools decreased vis-à-vis private schools. According to the results presented here, the increase in the number of teachers in the system, pushing the pupil-teacher ratio in Argentina to 11, had no effect on learning outcomes. The microsimulation further confirms that changes in the system's ability to transform inputs into outcomes accounted for most of the changes in test scores. Overall, the study shows the ineffectiveness of input-based education policies to improve learning outcomes in Argentina.

This paper is a product of the Education Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at rdehoyos@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Understanding the Trends in Learning Outcomes in Argentina, 2000 to 2012

Rafael de Hoyos
World Bank

Peter A. Holland
World Bank

Sara Troiano
Universitat Pompeu Fabra

JEL Classification: I20, I21, I22, O10

Keywords: Education, Quality, PISA, Determinants to learning, Argentina

Understanding the Trends in Learning Outcomes in Argentina, 2000 to 2012*

Rafael de Hoyos

Peter A. Holland

Sara Troiano

I. Introduction

Human capital is one of the most important determinants of productivity, the engine of long-term economic growth. As shown in the literature review by Hanushek and Wößmann (2007), years of schooling in itself is not conducive to human capital formation, productivity, and growth. Rather, what triggers a positive relationship between education and economic growth are the skills and abilities acquired in school. According to the authors, these abilities are well measured by standardized tests.

In addition to economic development, quality of education is increasingly acknowledged as crucial to increasing health outcomes, reducing poverty, and fostering civic participation. For instance, Cutler and Lleras-Muney (2008) estimate that the health returns to education may be even higher than the monetary returns. World Bank (2011) shows that more and better education leads to a citizenry that is more responsible, more able to cope with shocks, better at parenting, more able to sustain a livelihood, and better at adopting new technologies. All of these outcomes help drive progress for society at large.

In Argentina, as in other Latin American countries, learning outcomes are well below levels that their per capita GDP would predict, jeopardizing the country's long-term development potential. While other countries in the region have improved learning outcomes since 2000, measured by the OECD's PISA test, Argentina's scores show no progress (at best), or even a marginal decline between 2000 and 2012. Little is known about what is behind these disappointing levels and trends in test scores and, in particular, whether the increased resources that Argentina allocated to education during the period of analysis improved learning outcomes.

In 2002 Argentina suffered a deep economic crisis that contracted GDP by 20 percent, coinciding with a significant reduction in learning outcomes between 2000 and 2006. Output per person increased steadily between 2003 and 2012, along with a dramatic fall in inequality, as the Gini coefficient declined from 0.50 to 0.39 (SEDLAC, 2015). After the approval of the education finance law of 2006, Argentina increased its spending on education from 3.5 percent of GDP in 2005 to close to 6 percent in 2013¹ (Albornoz, 2015). Despite favorable economic performance and a massive increase in resources spent on education, test scores showed only a marginal improvement between 2006 and 2012. More importantly, it is unclear to what extent these trends in learning outcomes are explained by trends in variables that are out of the control of the education system (such as household incomes) and how much can be attributed to education policies. In addition, to our knowledge, there has been no attempt to quantify the impact of this recent increase in education sector financing on test scores.

*The authors are grateful for useful comments from Christian Bodewig. Authors can be contacted at rdehoyos@worldbank.org (Rafael de Hoyos), pholland@worldbank.org (Peter A. Holland) and troiano.sara@gmail.com (Sara Troiano).

¹These figures represent revised numbers based on INDEC's updated calculation for GDP in 2014.

This study examines the factors that contribute to student learning outcomes in Argentina, using data from PISA for years 2000, 2006, 2009, and 2012. We estimate a year-specific production function whose results are then used to explain the trends in learning outcomes in terms of changes in inputs and changes in parameters linked to each of them (the *efficiency effect*). The paper is organized as follows: Section II presents a brief review of the literature. Section III describes the trends in learning outcomes using four rounds of PISA: 2000, 2006, 2009, and 2012. Section IV presents the methodology, including the education production function and the microsimulation used to decompose the changes in learning outcomes in terms of inputs, parameters, and residuals. Section V presents the findings. Finally, Section VI offers reflections on implications for education policy in Argentina.

II. Literature Review

Learning is a complex, multi-faceted process. There are two types of factors that influence learning: those that are endogenous to the education system, that is, within the sphere of influence of education policy makers, and those that are exogenous, or beyond the sphere. This section cites some of the conventional exogenous factors identified by the literature and briefly summarizes the evidence on endogenous inputs, dividing them into two broad groups: school characteristics and institutional setting.

The evidence linking learning outcomes with exogenous factors such as socioeconomic background and individual characteristics is vast (Coleman et al., 1966, Lee and Barro, 1997, Fertig and Schmidt, 2002, and many others). There is some evidence that, in general, socioeconomic status is one of the most important determinants of learning outcomes (Mizala, Romaguera, and Urquiola, 2007, and others) but relatively less important than school factors in developing countries (Heyneman and Loxley, 1983, Fuller, 1987, Baker, Goesling, and Letendre, 2002, and others). With regard to parents' education, the literature also points to this factor as being correlated with student achievement and highly correlated with grade attainment in developing countries (Aturupane, Glewwe, and Wisniewski, 2007), particularly in the case of the mother's education (Glewwe and Jacoby, 1994, Glick, Randrianarisoa, and Sahn, 2011). There is overwhelming evidence that family background has a stronger effect on achievement in reading than in math and science (Fuchs and Wößmann, 2004).

Evidence from the United States shows that the home environment has a special impact in a child's early years, leading to student academic success later (Carneiro and Heckman, 2004). For developing countries, a longitudinal analysis of determinants of learning in 27 countries finds that homes that had more books available were correlated with children who performed better in school (Evans et al., 2010), a finding consistent with previous investigations of the presence of books in the house (Fuchs and Wößmann, 2004, and Yayan and Berberoglu, 2004). This is also corroborated by data from PISA, looking at 20 OECD countries, where students with more books in the home had an educational advantage over students with fewer (Thompson and Johnston, 2006). Other elements of the home environment that have seemed to impact cognitive outcomes in children are the existence of opportunities for learning, the warmth of mother-child interactions, and the physical conditions of the home (Brooks-Gunn and Duncan, 1997, and Majopribanks, 1994).

In terms of student characteristics, looking across all OECD countries, girls outperformed boys by 38 points (an average of one year) in reading (OECD, 2014). For math, boys outperformed girls by 11 points (an average of three months). Another student characteristic that is an important

determinant of learning is age for grade, a proxy indicator for students who may have repeated one or more years in school.

These findings linking learning outcomes with exogenous factors at the individual level have also been confirmed in Argentina. Recent evidence using standardized testing data (including PISA, LLECE, and Argentina's national standardized test, *Operativo Nacional de Evaluación, ONE*) have found that socioeconomic background and individual characteristics, such as attending preschool and not repeating grades, are hugely important in determining learning outcomes in students (Albornoz, Furman, Podesta, Razquin, and Warnes, 2015, Marchionni, Pinto, and Vazquez, 2013, Fresoli, Herrero, Giulioldoli, and Gertel, 2007, and Kruger, 2011).

2.1 School Characteristics

A recent meta review of studies and impact evaluations in developing countries finds strong evidence that textbooks and other similar materials increase student learning, as do basic furniture such as desks, tables, and chairs, and characteristics of the school buildings themselves. Importantly, there is some evidence that students in the bottom of the performance distribution are most affected by poor school conditions (Fertig and Schmidt, 2002). The availability of electricity also has a strong effect (Glewwe, Hanushek, Humpage, and Ravina, 2014). While the presence of computers on their own doesn't seem to generate an impact, computer-assisted learning programs have proven effective when instruction is integrated into the curriculum and tailored to individuals, and where teachers are well prepared to support the students (Evans and Popova, 2015). The availability of computer facilities and science laboratories is also significantly correlated with better learning outcomes in Mexico (De Hoyos, Espino, and García, 2012).

Evidence from Argentina is consistent with these findings. Educational resources such as libraries, multi-media resources, science laboratory equipment, facilities for the fine arts, and computers were all found to be correlated with higher test scores (Fuchs and Wößmann, 2004, and Santos, 2007), although school infrastructure was found to be less important (Santos, 2007).

Among the most potentially influential factors leading to learning are teachers and school directors. Recent work in the United States underlines both the tremendous impact teachers can have, as well as the great variance of the performance among teachers, and the lasting consequences this can have on students. Students lucky enough to have great teachers can gain 1.5 grade levels or more, while students with poor teachers can end up mastering only half or less of the curriculum (Hanushek and Rivkin, 2010). Research from developing countries also corroborates the singular potential impact of teachers: Glewwe et al. (2014) find that, of 63 studies estimating the effect of teacher experience, 43 show no statistically significant impact. But of the 20 that do show an impact, 17 of them are positive. The evidence for teacher education is slightly stronger, as is the evidence for teachers' knowledge of subject matter. As with computers, the evidence seems to indicate that effective interventions by teachers should be tailored to their skill level and provide them with specific guidance and support in their teaching (Evans and Popova, 2015).

Recent evidence from Latin America also points to the strong effect of the quality of teachers on learning outcomes. Using 2006 SERCE data, Bruns and Luque (2014) show that, within the same schools, large variations exist in student learning outcomes between classrooms, accounting for more than 40 percent of the total variation in Nicaragua and Cuba. Although these measures probably reflect policies of streaming children by ability into different classrooms, such large learning gaps are revealing of how being assigned to different teachers influences student

outcomes—even within the same school (Bruns and Luque, 2014). Evidence from Guatemala and Peru presented in Marshal and Sorto (2012) shows the importance of teachers’ content mastery: teachers who perform better on math tests have students who also perform better in that same subject. Using data for Mexico, De Hoyos, Espino, and García (2012) show that the proportions of teachers with a post-graduate degree and school directors assigned via a meritocratic process are both positively associated with higher test scores in math.

There is evidence from Argentina that also points to the importance of teachers within a context of school autonomy. Santos (2007) finds that giving teachers a higher degree of autonomy on decisions of budget allocation within schools, selection of textbooks, establishment of disciplinary policies, and involvement in student assessment policies is associated with better student test scores. According to Santos (2007), students performed better in schools where teachers had high expectations and where principals felt that there was a strong relationship between students and teachers. In contrast, Abdul-Hamid (2007) finds that teacher certification did not show a significant positive impact on learning outcomes. Finally, Fresoli et al. (2007) find that, among teachers’ years of education, their years of experience, and their amount of in-service training, only the number of years of education was significantly correlated with student learning (using data from LLECE). ONE data, meanwhile, shows a correlation between years of education as well as years of service (though only statistically significant in mathematics, not Spanish).

2.2 Institutional Setting

There are a number of elements related to institutional settings that could influence student performance: pupil-teacher ratios, length of the school day, degree of school autonomy, and involvement of parents, to name a few. These institutional factors represent about one-quarter of the between-country variations in performance on PISA 2000, and account for a large portion of students’ success (Fuchs and Wößmann, 2004, and Fertig and Schmidt, 2002). Fuchs and Wößmann (2004) find that students in systems that have external examinations or standardized tests perform better across all three subjects examined in PISA. Student performance seems to be higher when schools have more autonomy over decisions such as the hiring of teachers, textbook choice, and budget allocations within schools. The combination of these two elements is even more powerful: the performance effects of school autonomy are more beneficial in systems where external exit exams are in place, underscoring the idea that external exams are the “currency” of the school system (Fuchs and Wößmann, 2004). However, some of these findings seem to differ depending on a country’s level of development. While developed countries are able to translate increased autonomy into improved student learning, increased autonomy seems to harm student achievement in developing countries (Hanushek, Link, and Wößmann, 2013).

Among other institutional reforms in developing countries, pupil-teacher ratios (or class size in general) have been most widely studied, with very mixed findings (Hanushek, 2003, and Kruger, 2003). Teacher absenteeism and whether teachers assign homework are strong predictors of student success (Pelletier and Normore, 2007). Providing school meals is inconclusive, as is multi-grade teaching, while extending the school day shows more positive results (Glewwe et al., 2013).

Evidence suggests that longer school days do improve learning in Uruguay and Chile (Cerdan-Infantes and Vermeersch, 2007, and Bellei, 2009). There is some evidence that school autonomy in the form of school-based management (SBM) has resulted in gains in enrollment, reduced repetition and dropout, and raised parental participation (Gertler, Patrinos, and Rubio,

2012). Evidence of the effect of school-based management on student learning is more mixed and has generally employed weak evaluation techniques, though some recent research shows promising impact on Spanish in 3rd grade (Santibañez, Abreu-Lastra, and O'Donoghue, 2014).

In Argentina, the literature on the impact of the institutional setting on learning outcomes is also mixed. Regarding longer school days, one study in Buenos Aires shows some positive results on secondary school graduation rates (Llach, Adrogué and Gigaglia, 2009), but impacts on learning weren't estimated. Another study shows that schools in large cities had higher average test scores than other schools, and that top achieving schools were in large and medium-sized cities, while schools in villages and small towns had higher variation than schools in cities (Abdul-Hamid, 2007). Pupil-teacher ratios appear to have no effect (Fresoli et al., 2007).

One question that has garnered much attention in Argentina is whether the observed performance gap of private schools over public schools is attributable to the private nature of the school. Both Marchionni et al. (2013) and Albornoz et al. (2015), examining PISA 2009 and PISA 2009 and 2012 data respectively, conclude that when controlling for peer effects the performance advantage of private schools disappears. Formichella (2011), using data from PISA 2006, also finds that private school outperformance is in large part accounted for by having students who are more prepared to learn. The high segmentation of students, which is, to a lesser extent, also present in public schools, features prominently in the literature, and seems to be an important element behind the inequality of learning in Argentina (Kruger, 2011).

III. Country Context and Trends in Argentina's Learning Outcomes

3.1 The Education Sector in Argentina

Argentina was one of the first countries in the region to achieve universal primary education, and more recently saw rapid expansion of secondary education from 55 percent in 1999 to more than 75 percent in 2012 (UNESCO UIS, 2015). Today, the education system serves about 11 million students,² nearly 4 million of whom are enrolled at the secondary level. Among secondary school students, 39 percent are in private schools, a proportion unchanged since 2007 (DINIECE, 2015). Upper secondary school, covering grades 10, 11, and 12, consists of general and vocational tracks and serves students aged 15 to 18. The majority of these are enrolled in a general track, with only about 16 percent following a technical track (DINIECE, 2011). Therefore, among the students sitting for the PISA test, the vast majority are following a general track, with only a few having spent some months on a vocational track prior to testing.

At the primary level, where about 4.5 million students are attending school, private schools have seen their numbers increase from 29 percent of students in 2007 to 36 percent in 2014 (DINIECE, 2015). There's much speculation about what is behind this growth. Some analysts cite supply-driven financing policies (Narodowski, 2002, and Narodowski and Moschetti, 2015), while others focus on demand-side factors relating to family choice of schools. This includes proximity to a school, regardless of whether it is public or private (Gomez Schettini, 2007), or matters relating more to perceptions of the services offered (Cafiero, 2008), including better equipment or better quality of instruction. Still other factors could relate to issues of accountability in private schools, with teachers striking less, and schools therefore more likely to stay open, or simply perceptions of frequent teacher absenteeism in public schools (Tosoni and Natel, 2010).

²Not including adult education.

With regards to financing, the combination of high economic growth and high levels of investment as mandated by the education finance law of 2006 has led to unprecedented increases in resources for the education sector over the last decade. The increase in investment in education as a share of GDP from less than 3 percent in 2003 to close to 6 percent in 2013, paired with an average GDP growth rate of 6 percent over that same period (World Bank, 2014), has translated into a rise in public spending on education from about 500 million pesos to nearly 1.6 billion pesos (Albornoz, 2015). As with any education system, these resources have principally financed teacher salaries, followed by infrastructure and other capital investments such as information and communications technology (ICT) and other school equipment.

Concerning spending on teachers, there is evidence that this period witnessed both large increases in teacher numbers and teacher salaries. While the student base expanded by about 10 percent between 2003 and 2013 (DINIECE, 2004, and DINIECE, 2015), the recent teacher census revealed that the number of teachers expanded by more than 20 percent over roughly the same period (DINIECE, 2015). This has given Argentina a pupil-teacher ratio of 11, the lowest in Latin America after Cuba (OECD, 2013). With regards to teacher salaries, Bezem, Mezzadra, and Rivas (2012) find evidence that real salaries (in 2001 pesos) rose on average by 76 percent between 2003 (450 pesos) and 2012 (794 pesos). Still, this is less than the overall tripling of public resources for education, implying that the increased sectoral financing has meant a substantial rise in other schooling inputs. In particular, there was a large expansion in public school buildings constructed: 1,751, the most in half a century (Aboldornoz, 2015). The investment in ICT has also been impressive, representing as much as 50 percent of the infrastructure investment in the years 2011 and 2012 (Albornoz, 2015).

3.2 Trends in PISA Results

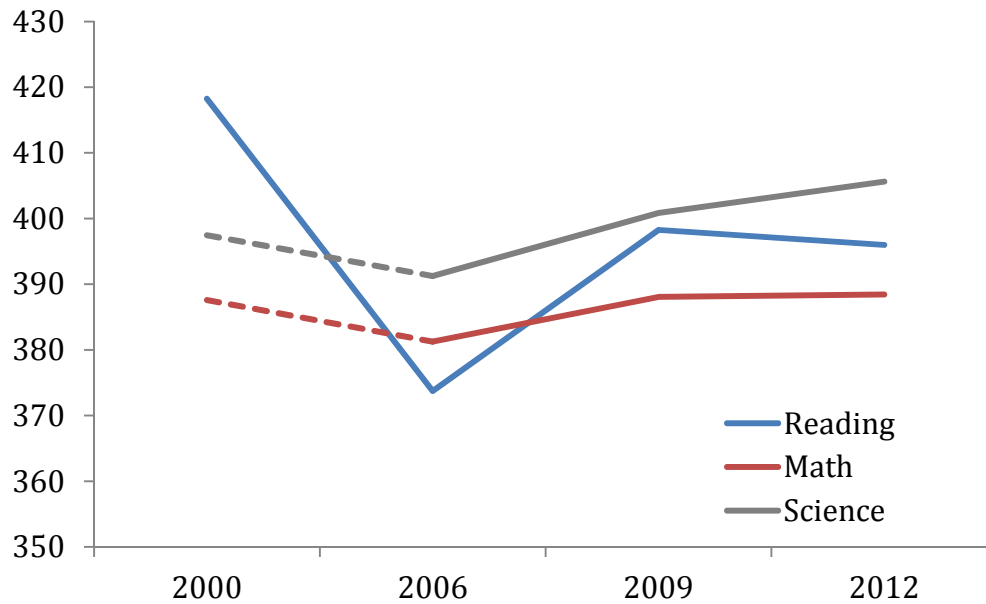
The OECD's Program for International Student Assessment (PISA) was first carried out in 2000, and since then a new edition has come out every three years. More than 70 countries have participated in PISA to date, making it the preferred dataset for cross-country comparisons. PISA is a standardized test assessing three subject areas: mathematics, reading, and science. It is not based on the curriculum of any particular education system. Rather, it evaluates students' ability to apply their knowledge and competences to problem solving. By construction, PISA had a mean of 500 and a standard deviation of 100 among OECD countries in 2000, in each subject area (see Annex A for more details on PISA's design).

This study uses micro data at the student level from 2000, 2006, 2009, and 2012, years in which Argentina participated in the PISA survey. PISA results are composed of a random sample of 15-year-olds who are attending school. The students are selected by a two-stage sampling strategy—first, by randomly selecting schools, and then by selecting 15-year-olds within the schools, also at random. In the case of Argentina, samples are representative at the national level only (DINIECE, 2004).³

³ A look at the sample suggests different degrees of representativeness in the editions considered. Using the survey's expansion factors, the weighted sample accounted in 2000 for more than 90 percent of the total 15-year-old population enrolled. Nevertheless, there is an important number of missing values. Fortunately the missing values are not for the learning outcome results but for variables capturing school characteristics and socioeconomic background where, in some cases, more than 40 percent of the observations report missing values in 2000. A first concern is the randomness of missing values. Intuitively, missing at random sounds like a very strong assumption. The missing

It is possible to identify three distinct phases in the evolution of PISA test score means over the period 2000-2012.⁴ As shown in Figure 1, mean scores decreased between 2000 and 2006 for all three subjects assessed in PISA, then increased between 2006 and 2009, and remained roughly stable between 2009 and 2012. Changes in reading scores are larger than those in math and science, with these last two areas showing very similar trends.

Figure 1: Trends in Learning Outcomes, 2000-12



Source: OECD's Program for International Student Assessment

Table 1 shows the changes in mean score by year and subject area, including the t-statistic for the test of constant levels of learning outcomes across any possible combination of years.⁵ A statistically significant negative change (at the 99 percent level) occurs in reading scores from 2000 to 2006 and a positive one from 2006 to 2009, while the change is not significant between 2009 and 2012. Changes in means in math and science are not statistically significant in any of the

values' proportion varies considerably across years. While a relatively large proportion of observations in 2000 report missing values, the share is almost insignificant in 2006 and even less so in 2009 but then again significant in 2012. Missing values are unlikely to be at random, particularly in the 2000 edition of the survey, as they seem to be related to students from the lowest quintiles of the income distribution. A second concern is that causes behind the missing values are unlikely to be the same across years. To limit potential bias, we exclude from the analysis variables that present more than 20 percent of missing variables in at least one of the four PISA rounds.

⁴ Note that, due to a change in the scaling methodology of test scores, only means in reading are strictly comparable throughout the period 2000-2012. Scores in math and science are comparable only starting from the 2003 and 2006 editions of the survey, respectively (OECD, 2010).

⁵ Final test scores in PISA are expressed in five plausible values (PV), rather than a point estimate like a weighted likelihood estimate (WLE), where the mean final score is the simple average of the five plausible values. The use of PVs, however, implies an additional source of variation known as imputation variance. Moreover, the arbitrary choice of items that ensure comparability of tests across PISA editions gives rise to a linking error that has to be taken into account when estimating the standard errors of the difference in means scores (OECD, 2010). The use of a t-statistic or similar analysis that does not take into account the PISA survey design and PV would lead to a serious underestimation of variance and misinterpretation of results (see Annex A for more details).

above combinations, although the gradual increase in both areas between 2006 and 2012 is statistically significant at conventional levels (95 percent) for math but not for science (90 percent only).

Table 1: Changes in Learning Outcomes, by Year and Subject Area

	2006	2009	2012
Reading			
2000	-44.53 (-3.41)	-19.99 (-1.68)	-22.27 (-1.85)
2006	-	24.54 (2.63)	22.26 (2.30)
2009	-	-	-2.28 (-0.36)
Mathematics			
2006	-	6.81 (0.89)	7.18 (2.30)
2009	-	-	4.79 (-0.36)
Science			
2006	-	9.59 (1.20)	14.39 (1.81)
2009	-	-	4.79 (0.76)

Source: Authors' calculations based on PISA-OECD. T-statistics included in parenthesis.

Although the evolution in average test scores is an important indicator of the quality of education services, equally important is the variation or dispersion of test scores. Given its linkages with future labor market outcomes, the inequalities that we observe today in test scores is a good predictor of the inequalities in incomes that will prevail in the future (Checchi and van de Werfhorst, 2014). Table 2 shows the evolution in dispersion in test scores for all three subjects, over the period 2000 to 2012. The dispersion measure presented in Table 2 is the coefficient of variation. Between 2000 and 2006, when learning outcomes in reading—the only subject allowing a comparison between those two years—registered a large decrease, the dispersion also increased substantially. The increase in dispersion in reading test scores was reversed during the period 2006 to 2012. A reduction in inequality of learning outcomes during the period 2006 to 2012 was also observed in math and science test scores. As shown below, the reduction in dispersion of test scores is largely accounted for by improvements in performance by the bottom end of the distribution.

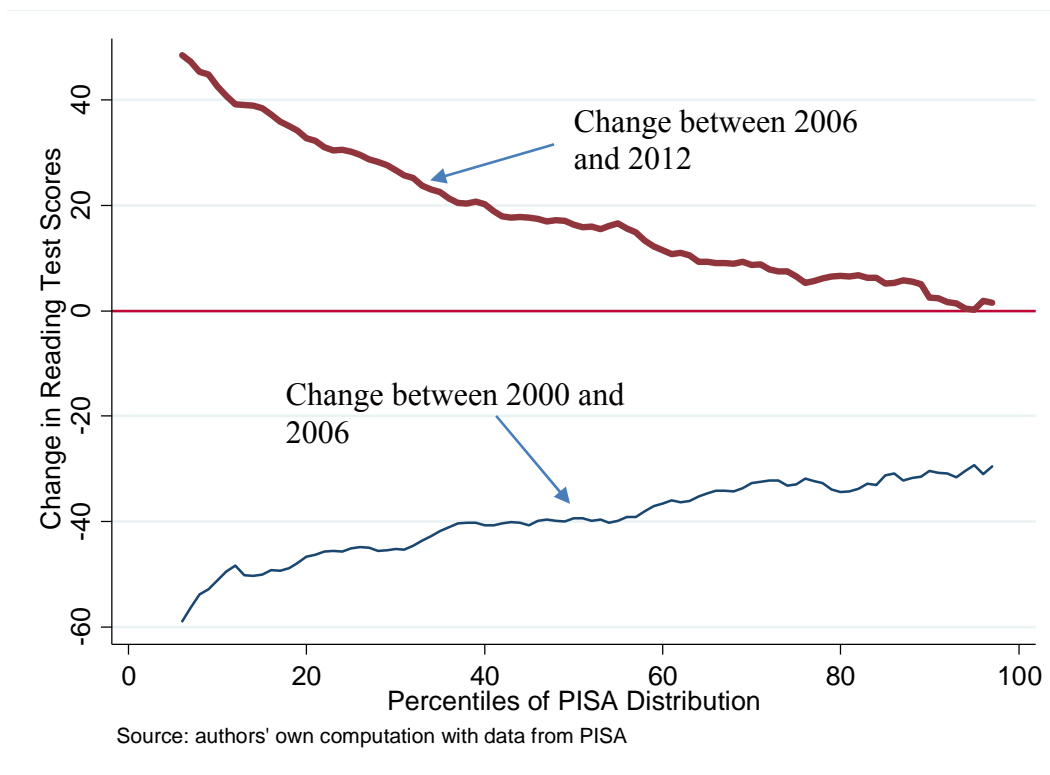
Table 2: Evolution of Dispersion in Learning Outcomes, by Year and Subject Area

	2000	2006	2009	2012
Reading	23.6	28.7	25.0	23.8
Math	28.1	23.2	21.9	19.5
Science	24.7	23.2	23.3	20.8

Source: Authors' calculations based on PISA/OECD. Dispersion is measured as the coefficient of variation defined as the standard deviation over the mean multiplied by 100.

Another way of illustrating the heterogeneity of changes in test scores along different parts of the distribution of learning outcomes is through a growth incidence curve (GIC). Originally developed for incidence analysis in the poverty and inequality literature, the GIC in the case of learning outcomes computes changes in average test scores by percentiles of the test scores distribution (for methodological details see Ravallion and Chen, 2003). Therefore, a GIC with a positive slope would indicate a regressive change in test scores, that is, students with initial lower scores improve less (or decrease more) than those with higher ones. The GIC summarizing the changes in reading test scores for the period 2000-12 is presented in Figure 2.⁶ Between 2000 and 2006, average test scores declined along the entire distribution of learning outcomes. However, the decline in reading test scores was substantially larger among students in the bottom 20 percent of the distribution of scores in 2006 vis-à-vis their peers in 2000. The decline in learning outcomes is lower as one moves to better-performing percentiles. The reverse is true for the period 2006-12. Except for students around the 95th percentile of the test score distribution, everybody did better in 2012 than in 2006. Not only did test scores increase during this period, but those that gained the most were the most disadvantaged students in the bottom part of the test score distribution.

Figure 2: Growth Incidence Curve for the Changes in PISA (Reading Test Scores)



IV. Methodology

Since the pioneering work of Coleman et al. (1966), researchers have used several methodological approaches to estimate what has been called the education production function. The aim of this approach is to link learning outcomes (outputs) with observable characteristics

⁶ Since PISA results are normalized, the y-axis in Figure 2 depicts the absolute change in PISA test scores as opposed to the percentage change as would be the case in a conventional GIC.

(inputs) that, *a priori*, could intervene in the development of cognitive abilities. As discussed in Section II, the set of inputs that are potentially associated with the formation of cognitive abilities includes individual, family, school characteristics, and institutional setting.

To formalize, let us define $Y_{i,t}$ as the learning outcome of student “ i ” in period “ t ” and $(X_{i,t}^1, X_{i,t}^2 \dots X_{i,t}^K)$ as the set or vector of all “ K ” inputs that are related with learning outcomes of student “ i ” in period “ t .” Therefore, learning outcomes can be defined by the following general expression:

(1)

$$Y_{i,t} = F(X_{i,t}^1, X_{i,t}^2 \dots X_{i,t}^K)$$

Function $F(\cdot)$ above captures the ability of the education system—or any other institution involved in the learning process—to transform inputs into learning outcomes. Notice that the technology defining $F(\cdot)$ is constant across students and over time, and that the function is general enough to allow for interaction terms and non-linearities. Although $F(\cdot)$ has been represented in many reduced forms, there is no consensus regarding the variables it should include nor its functional form. In the absence of structure to guide the definition of $F(\cdot)$, an assumption that the relationship between learning outcomes and past and present education inputs is linear has been the guiding principle (Todd and Wolpin, 2003):

(2)

$$Y_{i,t} = \alpha_t + \sum_{k=1}^K \beta_t^k X_{i,t}^k + \varepsilon_{i,t}$$

where $(\alpha_t, \beta_t^1 \dots \beta_t^K)$ are parameters to be estimated and $\varepsilon_{i,t}$ are random terms with an independent distribution approximating to the normal, zero mean, and known variance. Therefore:

(3)

$$E[Y_t|X_t] = \hat{\alpha}_t + \sum_{k=1}^K \hat{\beta}_t^k \bar{X}_t^k$$

Expression 3 can be used to explain differences on average learning outcomes between t and $t+I$ as a function of changes in estimated parameters $(\hat{\alpha}_t, \hat{\beta}_t^1 \dots \hat{\beta}_t^K) \rightarrow (\hat{\alpha}_{t+1}, \hat{\beta}_{t+1}^1 \dots \hat{\beta}_{t+1}^K)$ keeping inputs constant and changes in average inputs $(\bar{X}_t^1, \bar{X}_t^2 \dots \bar{X}_t^K) \rightarrow (\bar{X}_{t+1}^1, \bar{X}_{t+1}^2 \dots \bar{X}_{t+1}^K)$ keeping parameters constant.⁷

(4)

$$\begin{aligned} E[Y_{t+1}|X_{t+1}] - E[Y_t|X_t] \\ = \Delta(\hat{\alpha}_t, \hat{\beta}_t^1 \dots \hat{\beta}_t^K) * (\bar{X}_t^1, \bar{X}_t^2 \dots \bar{X}_t^K) + \Delta(\bar{X}_t^1, \bar{X}_t^2 \dots \bar{X}_t^K) * (\hat{\alpha}_t, \hat{\beta}_t^1 \dots \hat{\beta}_t^K) \end{aligned}$$

where Δ is the change operator. This simple decomposition would be enough to disentangle the effects of changes in, say, average socioeconomic status of students from changes in the importance of teachers to explain changes in learning outcomes between t and $t+I$.

⁷ In the inequality decomposition literature, expression (4) is better known as the Oaxaca-Blinder decomposition.

Expression 4 explains average changes in learning outcomes but tells us nothing about changes in different parts of the distribution or, in other words, changes in the dispersion of learning outcomes. Borrowing from the income inequality decomposition literature and following Bourguignon, Fournier, and Gurgand (2001) and Bourguignon, Ferreira, and Leite (2008), among others, we extend or generalize expression 4 to the whole distribution of learning outcomes. We define $I(\mathbf{Y})$ as an index measuring dispersion or inequality of the vector of learning outcomes \mathbf{Y} , like, for example, the decile ratio or the Gini coefficient. Based on Equation 2, $I(\mathbf{Y})$ can be defined as follows:

$$(5) \quad I(\mathbf{Y}_t) = I(\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t)$$

where $\mathbf{X}_t, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t$ are vector-matrix of inputs, parameters, and unobservables determining learning outcomes and its dispersion at t . Therefore the changes in $I(\mathbf{Y})$ between t and $t+1$ can be defined by the following expression:

$$(6) \quad \Delta I(\mathbf{Y}) = I(\mathbf{Y}_{t+1}) - I(\mathbf{Y}_t) = I(\Delta \mathbf{X}, \Delta \boldsymbol{\beta}, \Delta \boldsymbol{\varepsilon})$$

where $\Delta \mathbf{X}, \Delta \boldsymbol{\beta}, \Delta \boldsymbol{\varepsilon}$ capture the changes in vector-matrix of inputs, parameters, and unobservables between t and $t+1$. In other words, the change in average learning outcomes and its distribution can be decomposed by changes in the inputs available, the parameters attached to each input, and idiosyncratic random components.

4.1 Microsimulation Principle

So far, we have shown how to parameterize learning outcomes in order to identify the elements determining its level and distribution. The estimated parameters of Equation 3 can be used to perform microsimulation analysis to isolate the effect of changes in each of the three elements—or each of its subcomponents—on mean learning outcomes and its dispersion.

Once all the elements of Equation 6 are in place, we can create counterfactual experiments of nature asking *what would the distribution of learning outcomes look like had the elements of, say, \mathbf{X} been the only change occurring between t and $t+1$?* For example, let us say that the average socioeconomic status of students in Argentina changed between t and $t+1$ due to a negative macroeconomic shock and we would like to know how this shift affected mean learning outcome and its distribution. We can compute a hypothetical distribution of learning outcomes where the only element in (6) that is changing is socioeconomic status:

$$(7) \quad I(\mathbf{Y}_t^l) = I(\mathbf{X}_t^l, \boldsymbol{\beta}_t, \boldsymbol{\varepsilon}_t)$$

where \mathbf{X}_t^l contains the imputed value of socioeconomic status observed in $t+1$: X_{t+1}^k . $I(\mathbf{Y}_t^l)$ is a *simulated*, unobserved, hypothetical inequality or dispersion index where learning outcomes of each student in the database are allowed to change as a result of the change in X_{t+1}^k , and all other elements are kept fixed. This type of counterfactual exercise is quite powerful, since it enables us to identify the quantitative effect of a change in each element defining Equation 6: parameters, covariates, and residuals.

V. Results

5.1 Regression Results

We analyze the contribution of different variables to student learning outcomes according to the production functions presented in the previous section distinguishing between: (1) exogenous factors, i.e. student characteristics and socioeconomic background, and (2) endogenous factors, i.e. school characteristics (infrastructure and pedagogical materials, teacher, and school director characteristics), and factors linked to the institutional setting.

Table 3 shows the year-specific mean of all the variables included in the estimation of the education production function. Student characteristics include gender and a dummy variable equal to one when the student is not in 10th grade at the time of the test as a proxy for late enrollment or repetition. Socioeconomic background includes a family wealth index,⁸ education level of father and mother, and the availability of books at home. School characteristics include whether a school is public or private, the school location (rural versus urban), school size in terms of total number of students, student-teacher ratio, the share of teachers who completed tertiary education, and the number of Internet-connected computers per student. Finally, the institutional setting variables capture the degree to which the school has authority to fire teachers, formulate school budgets, decide on budget allocations, establish disciplinary policies, establish student assessment policies, determine courses, and decide which courses to offer.

⁸ The index of family wealth is based on the students' responses on whether they had the following at home: a room of their own, a link to the Internet, a dishwasher (treated as a country-specific item), a DVD player, and three other country-specific items; and their responses on the number of cellular phones, televisions, computers, cars, and rooms with a bath or shower.

Table 3: Average value of the variables included in the estimation of the production function

Variable	2000	2006	2009	2012
Student's characteristics				
Gender (female)	0.56	0.53	0.54	0.52
Attending at least 10th grade (not over age)	0.72	0.72	0.64	0.64
Socioeconomic background				
Wealth index	-0.90	-1.27	-0.95	-0.90
Mother's highest level of education				
<i>Did not complete primary school</i>	3%	12%	11%	8%
<i>Complete primary school</i>	37%	25%	22%	19%
<i>Complete uppersecondary school</i>	17%	15%	12%	20%
<i>Complete lower secondary school</i>	24%	18%	18%	22%
<i>Complete tertiary</i>	19%	30%	37%	31%
Father's highest level of education				
<i>Did not complete primary school</i>	3%	12%	12%	9%
<i>Complete primary school</i>	38%	24%	24%	22%
<i>Complete uppersecondary school</i>	19%	16%	12%	21%
<i>Complete lower secondary school</i>	20%	21%	20%	22%
<i>Complete tertiary</i>	20%	26%	31%	26%
Books at home				
<i>0-10</i>	31%	28%	28%	34%
<i>11-100</i>	31%	54%	53%	52%
<i>more than 100</i>	38%	19%	18%	14%
School characteristics				
Public school	0.58	0.61	0.62	0.65
Rural area	0.24	0.26	0.29	0.25
School size				
<i>Up to 200 students</i>	18%	21%	14%	21%
<i>200-500 students</i>	40%	39%	41%	38%
<i>500-1000 students</i>	25%	25%	31%	33%
<i>> 1000 students</i>	18%	15%	14%	9%
Student-teacher ratio (full time + 0.5 part-time)	8.44	11.36	12.00	9.12
Proportion of teachers with tertiary education	0.06	0.16	0.11	0.16
Computer with web connection available per student	0.01	0.02	0.03	0.06
Shortage/Inadequacy of instructional material (e.g. textbooks)	0.27	0.28	0.39	0.40
Institutional setting				
Autonomy - firing teachers	0.35	0.37	0.39	0.41
Autonomy - formulating the school budget	0.33	0.33	0.37	0.35
Autonomy - deciding on budget allocations	0.49	0.62	0.72	0.53
Autonomy - establishing student disciplinary policies	0.92	0.69	0.72	0.80

Autonomy - establishing student assessment	0.84	0.44	0.77	0.76
Autonomy - determining courses content	0.73	0.41	0.29	0.32
Autonomy - deciding which courses are offered	0.73	0.49	0.11	0.15

Notes: authors' own computation with data from PISA. All means take into account PISA survey design (see Annex A for details)

The selection of variables to be included in the estimation of the production function was restricted to those which were included in all four rounds of PISA. Some variables that were included in the four rounds were excluded from the estimation when 20 percent or more of the observations had missing values. The parameters of the production function were estimated with a variance-covariance matrix accounting for the additional variation in PISA results due to the plausible values methodology followed in the test design (see Annex A for details).⁹ All the estimations and subsequent simulations are based on reading learning outcomes since it is the only subject area for which a 2000-2012 comparison is valid.

The estimated results are presented in Table 4. The proportion of explained variance by the production function (as captured by the R^2) is relatively high at around 40 percent, a level close to the one reported in Santos (2007). The model's goodness of fit varies little across years. This might be the outcome of restricting the estimation of the production function to the same variables across surveys and using only those variables that have relatively low levels of missing values (less than 20 percent).

As expected, student characteristics and socioeconomic background variables are important determinants of student learning outcomes. In all four years, girls get significantly better reading results than boys, averaging close to a quarter of a standard deviation, which confirms the international evidence on gender differences in learning outcomes (Guiso et al., 2008). Grade-for-age is also significantly correlated with learning outcomes across all years and with very large effects: 15-year-old test takers who are not yet in grade 10 (the accurate grade-for-age), capturing either late entry into the education system or, more likely, repetition of grades, achieve learning outcomes that are close to three quarters of a standard deviation lower than 15-year-olds who attend 10th grade.

In terms of family background, mother's education is correlated with better results but only in 2009 and 2012, and father's education is not statistically significant in most years. In both 2006 and 2009, family wealth is a significant predictor, but insignificant in 2000 and 2012. With regards to household factors, having books at home is a consistently important predictor of students' performance, in line with many other studies in the literature. There is a substantial degree of multicollinearity between the number of books at home, wealth index, and parents' education. Therefore it is not surprising that, contrary to what was expected *ex-ante*, the wealth index or mothers' education is for some years not statistically significant. We do not attempt to address the issue of multicollinearity, since the variables included in students' characteristics and socioeconomic background are simple controls of exogenous inputs, whereas the paper's main focus lies on the effects on learning of school characteristics and institutional settings.

⁹ The parameters of the production function were estimated in Stata using the user-written command *pv* developed by MacDonald (2008, revised in 2014).

Table 4: Education Production Function Estimation Results, Full Model

Independent Variables	2000	2006	2009	2012
Student's characteristics				
Gender (female)	23.024***	37.484***	23.617***	24.964***
Not enrolled in 10th grade (over age)	-75.138***	-74.123***	-64.849***	-58.565***
Socioeconomic background				
Wealth index	-2.694	15.153***	9.679***	1.399
Mother's highest level of education				
<i>Complete primary school</i>	-23.147*	-13.503	5.973	16.477**
<i>Complete upper secondary school</i>	-17.029	-0.463	20.580**	16.008**
<i>Complete lower secondary school</i>	1.07	-0.514	17.348***	28.020***
<i>Complete tertiary</i>	-5.317	2.291	31.514***	41.153***
Father's highest level of education				
<i>Complete primary school</i>	-17.789	7.318	5.466	6.889
<i>Complete upper secondary school</i>	-21.415	18.147	11.968*	9.34
<i>Complete lower secondary school</i>	-7.095	6.68	7.055	8.502
<i>Complete tertiary</i>	-13.847	21.692**	9.905	6.24
Books at home				
<i>11-100</i>	17.646***	17.220***	14.957***	20.237***
<i>more than 100</i>	36.722***	34.059***	34.712***	38.043***
School characteristics				
Public school	36.116**	-57.437	-12.59	-51.617***
Rural area	-6.966	-2.971	-25.010**	9.147
School size				
<i>200-500 students</i>	-14.008	44.646***	28.698**	0.586
<i>500-1,000 students</i>	4.244	53.530***	37.714***	3.688
<i>> 1,000 students</i>	-1.246	62.641***	33.401**	4.623
Student-teacher ratio (full time + part-time)	0.906	-0.934	-0.242	0.074
Percentage of teachers with tertiary education	0.425	0.108	.426*	.351**
Computer with web connection, per student (x 100)	2.621*	1.456	3.297***	.598*
Shortage/Inadequacy of instructional material	-24.078***	-25.621**	-1.225	-8.827
Institutional setting				
Autonomy—firing teachers	52.721***	31.998	8.979	-0.597
Autonomy—formulating the school budget	8.709	-51.101***	16.39	-18.889**
Autonomy—deciding on budget allocations	2.416	1.083	1.739	13.313
Autonomy—establishing student disciplinary policies	-0.126	-2.497	-7.431	1.164
Autonomy—establishing student assessment	-8.651	-23.714*	9.612	-0.462
Autonomy—determining courses' content	25.107**	8.181	7.375	2.783
Autonomy—deciding which courses to offer	-4.022	4.037	-5.277	8.891

Constant	386.424***	402.407***	349.204***	385.594***
N	1,892	2,834	2,865	3,373
R2	0.38	0.39	0.43	0.37

Notes: All estimation results are accounting for PISA's survey design as explained in Annex A. As usual, * p<.10, ** p<.05, *** p<.01. The references category for mother's and father's education level is unfinished primary; the reference category for number of books at home is less than 11; and the reference category for school size is less than 200 students.

As previously discussed in this paper, the share of students enrolled in private schools in Argentina has increased constantly since 2003, reaching 39 percent of total students in secondary school and 29 percent in primary and secondary school in 2014 (DINIECE, 2015). One of several potential explanations behind this trend could relate to significant differences in the quality of education services between private and public providers. In an alternative specification that did not include controls for institutional settings, public schools were found to be significantly and negatively correlated with student performance across all years, even when controlling for students' individual and background characteristics (the results of this alternative specification are presented in Annex B). However, once we control for school autonomy variables, specifically those related to schools' authority to fire teachers, determine courses, and formulate the school budget, the negative effects associated with public school vanish. The results presented in Table 4 show that, once controlling for differences in institutional settings, public schools prior to 2012 were not different than private ones, at least in their ability to produce learning outcomes. This suggests that whether a school is public or private is not, in itself, determinant of school performance. What really matters is the degree to which those schools have the necessary autonomy to implement education policies. Interestingly, in 2012, even controlling for institutional setting, students in public schools achieved substantially lower learning outcomes, putting them behind their peers in private schools by almost half a standard deviation.

With the exception of 2009, once controlling for individual and socioeconomic background, rural schools seem to be no different than urban ones in terms of learning outcomes. This is somewhat surprising, since rural schools tend to be smaller, putting them at a disadvantage versus larger schools, which are much more predominant in urban centers, especially in 2006 and 2009. In those two years, schools with 200 students or more had reading test scores between 0.25 and 0.6 standard deviations higher than schools with fewer than 200 students.

Four variables are of particular interest since they could capture—albeit somewhat imprecisely—part of the increase in inputs brought about by the education finance law of 2006. These variables are: (1) the availability of computers with connectivity, (2) the student-teacher ratio, (3) the proportion of teachers with completed tertiary education, and (4) the school director's perception of whether instructional materials such as textbooks are adequately available. According to our results, a higher availability of a computer that is connected to the Internet—via a computer lab in the school—is positively related with higher reading test scores. However, by 2012, this is only significant at the 10 percent confidence level. Perhaps one of the most important results in our estimations is that the number of students per teacher has no effect on test scores in any of the PISA rounds. This has important budget and policy implications, since Argentina has a pupil-teacher ratio of 11, the lowest in the region and well below countries such as Chile (22),

Colombia (25), Brazil (28), and Mexico (30).¹⁰ On the contrary, the quality of teachers, as proxied by the share with tertiary education, has a positive and significant effect on learning outcomes. In 2000 and 2006, the inadequate provision of instructional material (as perceived by school directors) had a negative effect on learning outcomes. However, the effect basically goes to zero in 2009 and 2012, suggesting that any additional investment in this area after 2006 was ineffective in improving the quality of education services, assuming that school directors' perception captures the adequate provision of instructional material.¹¹

The learning outcomes effects of the school autonomy variables capturing the institutional setting show an overall negative trend. Two of the seven variables, authority to fire teachers and authority to determine courses, had a positive and significant effect in 2000 but zero effect in subsequent years. In 2006, two autonomy variables that were not previously significant showed a negative correlation with learning outcomes: having the legal authority to formulate the school budget and authority to set student assessment methods.

To further test this apparent negative change in the correlation between school autonomy and learning outcomes during the period 2000 to 2006, we run an alternative specification following Hanushek, Link, and Wößmann (2013). The authors aggregate the school autonomy variables into the following four indices: (1) autonomy in deciding teachers' salaries, (2) autonomy in budget and personnel decisions, capturing the formulation of school budget, allocation of school budget and firing of teachers, (3) autonomy on academic content, capturing course content and courses offered, and (4) autonomy in monitoring students' performance, capturing student disciplinary measures and assessments. The results of these alternative specification are presented in Annex C. Consistent with the results from the disaggregated version of school autonomy (Table 4), the alternative specification shows a negative change, between 2000 and 2006, in the correlation between two measures of autonomy and learning outcomes: budget and personnel decisions and academic content. While in 2000 schools that had more autonomy in these measures showed significantly better learning outcomes, this was no longer true in subsequent years. These are particularly important results suggesting that there was a reduction in the effectiveness of Argentine schools to transform more legal authority into better results.¹²

5.2 *Microsimulation*

We use the microsimulation techniques described in Section IV to decompose the changes in learning outcomes in Argentina measured by the PISA reading results of 2000, 2006, 2009, and 2012. The exercise identifies the relative importance of changes in the elements defining the production function to explain changes in mean learning outcomes and its distribution. To restrict the number of microsimulations presented in this section, we focus on the explanation behind two distinctive trends in reading learning outcomes as described in Section III, i.e. the decline between 2000 and 2006 and the subsequent partial recovery between 2006 and 2012. We create counterfactuals to decompose these two trends to the following three factors:

¹⁰ OECD (2013)

¹¹ An alternative explanation of these results is that in 2000 and 2006, the perception of school directors was capturing legitimate, objective, shortages of learning material. By 2009, after a significant increase in resources spent on education inputs, the perception of directors was capturing relative shortages of materials, i.e. in comparison with what other schools were getting, which is not necessarily correlated with learning outcomes.

¹² Notice that while overall autonomy in budget and personnel decisions did not change significantly between 2000 and 2006, the share of schools that decided on academic content had a substantial reduction (Table 3).

1. *Efficiency effect*: changes in the systems' capacity to transform inputs into learning outcomes captured by changes in the parameters of the education production function as presented in Table 4.
2. *Inputs effect*: changes in inputs in the production function, including levels and distribution of variables capturing individual characteristics, socioeconomic background, school characteristics, and institutional setting.
3. *Residual effect*: changes in the variance of the distribution of unobservables or the residual of the production function.

To construct the counterfactuals, the starting point is the estimated parameters of the production function presented in Table 4, which are used to generate an estimated residual for each individual in our sample ($\hat{\epsilon}_{i,t}$). To simulate the *ceteris paribus* effect of the changes in parameters of the production function (henceforth the *efficiency effect*) between 2000 and 2006, we take the estimated parameters for 2006 ($\hat{\beta}_{2006}$) and use them to score the set of inputs observed in 2000 ($\mathbf{X}_{t,2000}$) and then add the estimated residuals for that same year ($\hat{\epsilon}_{i,2000}$). The resulting distribution of learning outcomes is a hypothetical one capturing a scenario where the only change between 2000 and 2006 are the parameters of the production function or the *efficiency* of the system to transform inputs into reading test scores. In other words, the *efficiency effect* counterfactual answers the following question: how would test scores in Argentina have evolved between 2000 and 2006 had the efficiency of the system been the only change during that period?

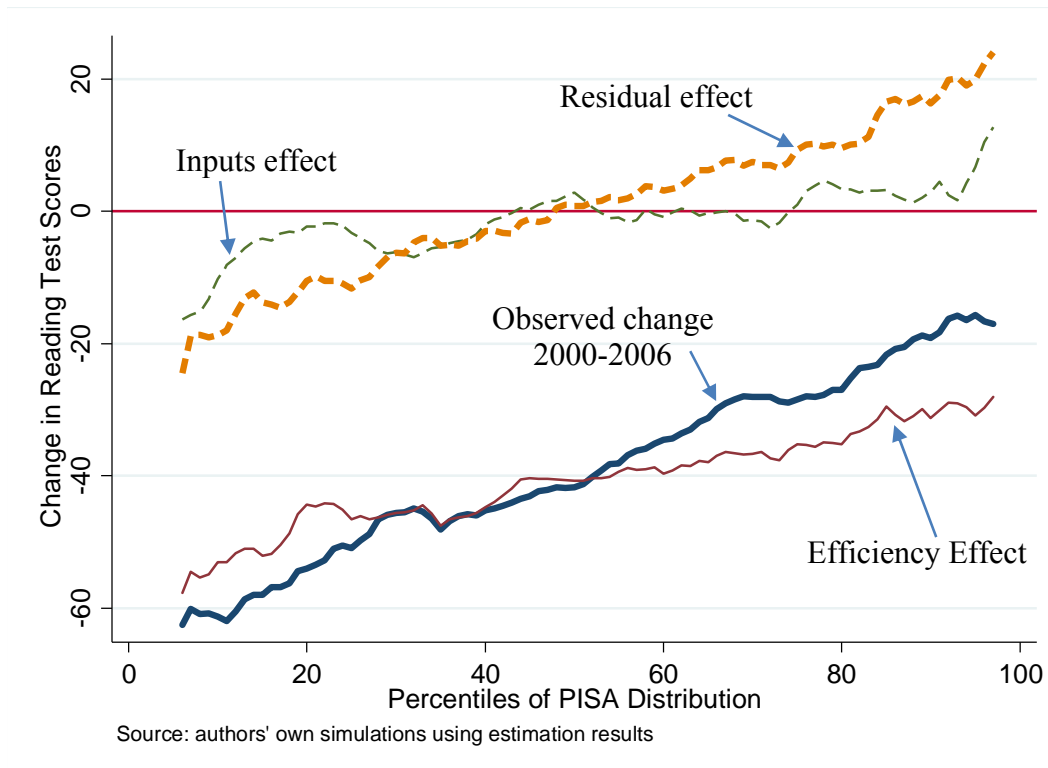
The *residual effect* is related to the errors of the production function. The residuals capture all inputs affecting learning outcomes that are not taken into account by the variables included in the regression. In the particular case of our estimated production function, things such as past inputs, students' efforts, quality of teaching, infrastructure, and school climate (among many others) would be left to the residual. Changes in the distribution of residuals can affect the distribution of learning outcomes. By construction, the residuals have mean zero in each year. However, the distribution of residuals does change over time. For instance, between 2000 and 2006 the standard deviation of the estimated residuals passed from 69 points to 85 points of PISA. To quantify the test score effects of this increase in the dispersion of residuals, following Bourguignon, Fournier, and Gurgand (2001), we modify the residuals in 2000 by the ratio of the standard deviation of residuals in 2006 over the standard deviation of residuals in 2000. In other words, the *residual effect* counterfactual answers the following question: how would test scores in Argentina have evolved between 2000 and 2006 had the change in the dispersion of the residuals been the only difference during that period?

Finally, the counterfactual capturing the *inputs effect* is the difference between observed test scores and the simulated test scores incorporating, simultaneously, the *efficiency* and *residual effects*. Notice that, since PISA is a repeated cross section, there is no way of creating a counterfactual where the only changes are the changes in inputs. While this counterfactual could be easily created if we were interested in decomposing changes in test scores at the mean (like in Oaxaca, 1973, or Blinder, 1973), there is no straightforward strategy for doing this when the purpose is to simulate changes along the entire distribution of test scores.¹³ The *input effect*

¹³ For a detailed discussion of the use of microsimulations to construct synthetic panels, see Bourguignon (2011).

counterfactual therefore answers the question: how would test scores in Argentina have evolved between 2000 and 2006 if the only change during that period were the education inputs?

Figure 3: Growth Incidence Curve Capturing the Microsimulation Results, 2000-2006



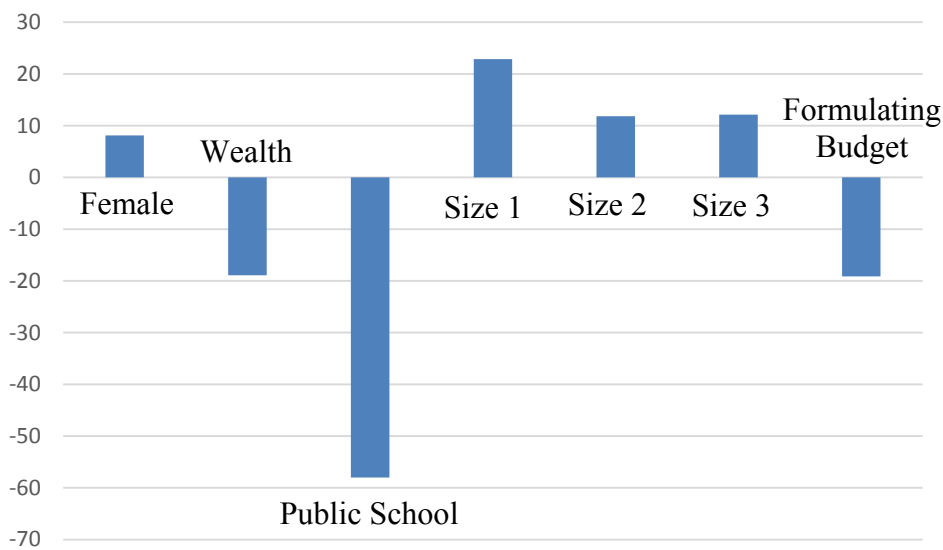
The results of these three microsimulations decomposing the changes in reading learning outcomes between 2000 and 2006 are summarized in Figure 3. All three effects are regressive in the sense that larger negative effects are simulated among students in the lower part of the distribution of PISA results. What the microsimulation uncovers is that most of the actual change in learning outcomes between 2000 and 2006 is accounted for by the *efficiency effect*. During the years following the crisis, the system experienced a substantial decrease in its ability to transform inputs into learning outcomes, which explains much of the decline in test scores between 2000 and 2006.

Figure 3 also shows that the changes in residuals' distribution had a regressive effect, reducing learning outcomes among poor performers by around 20 PISA points and increasing it by around the same amount among those performing relatively well. According to the simulation results, changes in inputs did not have a substantial impact on learning outcomes.¹⁴ The increase in the number of teachers (or a decrease in student-teacher ratio), the increase in the share of

¹⁴ The relatively low importance of the *inputs effect* may be surprising given the harsh crisis that hit Argentina in 2001-02. However, this can be explained by looking at the strong recovery between 2003 and 2006. By 2006, real income per capita was 24 percent higher than its 2000 level.

teachers with tertiary education, and the increase in the availability of computers (see Table 3 for a description of changes in inputs) had almost no effects on reading learning outcomes. This result is consistent with an abundant literature showing the failure of input-based education policies to improve learning outcomes (see Hanushek, 2003). In order to better understand how each independent variable contributed to a change in test scores, Figure 4 shows the isolated learning outcome effects of changes in those parameters with statistically significant changes between 2000 and 2006.¹⁵ The results show that there was a substantial reduction in the ability to produce learning outcomes in public schools and the parameters transforming school autonomy into test scores.¹⁶ For instance, if the parameter for public school in the education production function was the only element changing between 2000 and 2006, then PISA reading results in Argentina would have massively decreased by 57 points. Overall, the results presented in Figure 4 show that the decline in the system’s efficiency—and hence the decline in learning outcomes estimated between 2000 and 2006—is to a great extent explained by changes in a few parameters related to school characteristics (public versus private schools) and institutional setting (ability to formulate a budget).

Figure 4: Isolated PISA Effect of Changes in Selected Parameters in the Production Function, 2000-2006



Source: Authors’ simulations using estimation results.

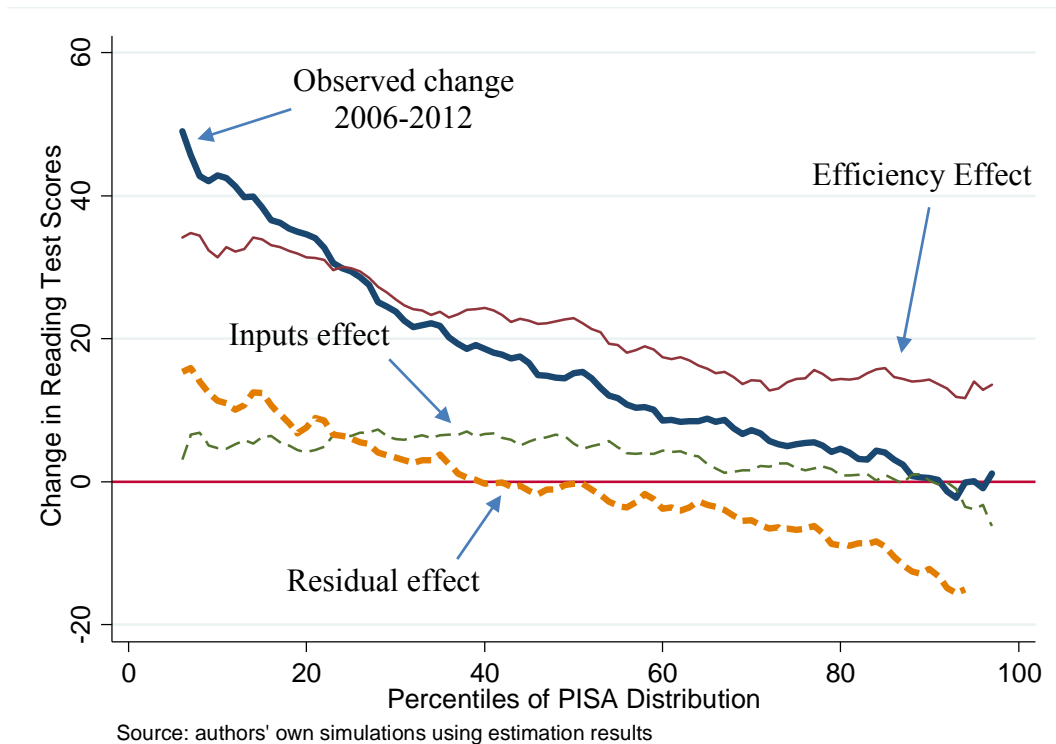
The microsimulation results decomposing changes in reading learning outcomes during the recovery period 2006-2012 are presented in Figure 5. The ineffectiveness of input-based education

¹⁵ The identification of parameters that changed significantly between years is done via a cross-model or cross-equation hypothesis testing using seeming unrelated estimations (SUEST).

¹⁶ The movement of students with relatively better backgrounds out of public and into private schools could account for a reduction in average learning outcomes among students in public schools. If this composition effect was at work, the share of total variance in test scores accounted for by the average difference between schools would increase over time. However, this is not the case. In 2000, differences between schools accounted for 51 percent of the total dispersion in test scores, while the same share was 49 percent in 2006.

policies in Argentina is confirmed by the results simulating the *inputs effect* in Figure 5. Had inputs been the only element changing in the production function between 2006 and 2012, then average test scores would have increased only marginally during that period.

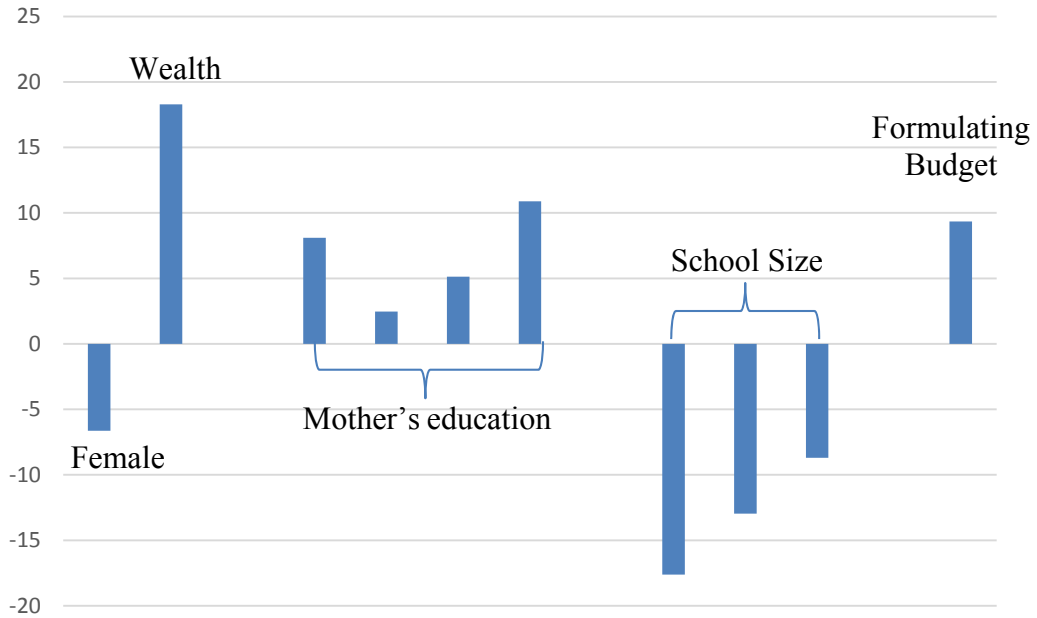
Figure 5: Growth Incidence Curve Capturing the Microsimulation Results, 2006-2012



As was the case for the period characterized by a decline in learning outcomes, the recovery period is explained, to a large extent, by the efficiency effect or the changes in parameters in the production function. Nevertheless, as shown in Figure 6, the set of parameters that change favorably during the recovery period are those related to students' characteristics and socioeconomic background. For instance, had the only element changing in the production function been the wealth parameter, it would have been enough to explain an increase of 18 points in learning outcomes between 2006 and 2012. The changes in the parameters for mothers' education are also contributing to the recovery during that same period.

In terms of the change in parameters of school characteristics and institutional setting during the period 2006 to 2012, learning outcome effect of public schools remains lower than in private schools despite the overall gains in reading test scores during those years. The system's effectiveness in transforming educational inputs into results improved, though only marginally, in areas such as the proportion of teachers with completed tertiary education and, to a lesser extent, the decision making at the school level (formulating school budgets).

Figure 6: Isolated PISA Effect of Changes in Selected Parameters in the Production Function, 2006-2012



Source: Authors' simulations using estimation results.

VI. Conclusions

Argentina's education system, which has shown virtually no improvements in basic cognitive abilities as measured by PISA test scores in the last 15 years, will have long-lasting negative effects on the country's productivity and economic growth prospects. Moreover, since the learning gap between students in public versus private schools increased during the same period, the education system's ability to reduce income inequality and promote social mobility is compromised. It is therefore central to Argentina's development agenda to understand the causes behind these unwelcome trends in learning outcomes.

Despite an important increase in the public resources invested in education inputs, such as rises in teacher numbers, teacher salaries, school infrastructure, and laptops, none of it seems to have significantly improved learning outcomes. Our simulation shows that the trends in PISA learning outcomes are explained not by the amount of inputs in the system but by how efficiently those inputs are used in the production of learning. This reinforces the finding from our regression analysis that quantity of teachers (as measured by pupil-teacher ratios) does not correlate with learning, and that other inputs (such as computers and instructional materials) seem to have marginal returns and diminish over time. In fact, the decline in learning outcomes between 2000 and 2006 is largely explained by the drop in test scores of students in public schools, while much of the recovery between 2006 and 2012 is accounted for by improvements in socioeconomic variables such as wealth and mothers' education levels.

The results presented here highlight the failure of input-based education policies to improve learning outcomes. This is not to say that education inputs are not important in the learning process, but their impact is limited and seems to diminish over time because it is subject

to the system's poor ability to use them in an effective way. Going forward, Argentine policy makers would do well to focus on those policy dimensions that have the most potential to transform existing inputs into learning. First among these would be a student assessment system that brings learning into sharper focus, and helps teachers, schools, supervisors, and other actors rally around the goal of improving learning. Monitoring and evaluation systems that provide schools with timely information, perhaps paired with school finance systems that reward performance, would also help align stakeholders behind the ultimate objective for Argentine schoolchildren. Finally, no reforms will succeed unless teachers have both a challenging and a rewarding career path that offers them opportunities for advancement and support in their efforts for excellence, and recognizes their performance as the most important part of the education system.

References

- Abdul-Hamid, H. 2007. "Assessing Argentina's Preparedness for the Knowledge Economy: Measuring Student Knowledge and Skills in Reading, Mathematical and Scientific Literacy with Evidence from PISA 2000." *Well-being and Social Policy*, Vol. 3(2): 41-66.
- Albornoz, F. 2015. "Education Public Investment, Schooling Quality, and Economic Growth: The Case of Argentina." World Bank background paper. August 2015.
- Albornoz, F., M. Furman, M.E. Podesta, P. Razquin, and P. Warnes. 2015. "*Diferencias educativas entre escuelas privadas y públicas en Argentina*". Unpublished working paper.
- Andrabi, T., J. Das, and A. Khwaja. 2009. "Report Cards: The Impact of Providing School and Child Test-Scores on Educational Markets." BREAD Working Paper No. 226, Bureau for Research and Economic Analysis of Development.
- Aturupane, H., P. Glewwe, and S. Wisniewski. 2013. "The Impact of School Quality, Socioeconomic Factors and Child Health on Students' Academic Performance: Evidence from Sri Lankan Primary Schools." *Education Economics*. 2013;21(1):2-37.
- Baker, D.P., B. Goesling, and G.K. Letendre. 2002. "Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the 'Heyneman-Loxley Effect' on Mathematics and Science Achievement." *Comparative Education Review*, 46, 291-312
- Banerjee, A., P. Glewwe, S. Powers, and M. Wasserman. 2013. "Expanding Access and Increasing Student Learning in Post-Primary Education in Developing Countries: A Review of the Evidence." Post-Primary Education Initiative Review Paper, Abdul Latif Jameel Poverty Action Lab, Cambridge, MA.
- Banerjee, A., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122(3), 1235-1264.
- Barrera-Orsorio, F., and L. Linden. 2009. "The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia." Policy Research Working Paper No. 4836, World Bank: Washington, DC.
- Bellei, C. 2009. "Does Lengthening the School Day Increase Students' Academic Achievement? Results from a Natural Experiment in Chile." *Economics of Education Review*, 629-640.
- Bezem, P., F. Mezzadra, and A. Rivas. 2012. "*Monitoreo de la Ley de Financiamiento Educativo*." Informe Final CIPPEC. Buenos Aires.
- Bourguignon, F. 2011. "Non-Anonymous Growth Incidence Curves, Income Mobility and Social Welfare Dominance." *Journal of Economic Inequality*, Springer, vol. 9(4), pages 605-627, December.

- Bourguignon, F., F. Ferreira and P. Leite. 2008. "Beyond Oaxaca–Blinder: Accounting for Differences in Household Income Distributions." *Journal of Economic Inequality*, Springer, vol. 6(2), pages 117-148, June.
- Bourguignon, F., M. Fournier, and M. Gurgand. 2001. "Fast Development with a Stable Income Distribution: Taiwan, 1979-94," *Review of Income and Wealth*, International Association for Research in Income and Wealth, vol. 47(2), pages 139-63, June.
- Brooks-Gunn, J., and G. Duncan. 1997. "The Effects of Poverty on Children and Youth." *The Future of Children*, 7, 55-71. Princeton University, NJ.
- Bruns, B., D. Filmer, and H.A. Patrinos. 2011. "Making Schools Work: New Evidence on Accountability Reforms." The World Bank: Washington, DC.
- Bruns, B. and J. Luque. 2014. "Great Teachers: How to Raise Student Learning in Latin America and the Caribbean." The World Bank: Washington, DC.
- Cafiero, M. 2008. "*Experiencias Escolares en Familias de Sectores Populares. Fragilidades y Valoraciones de la Escuela en las Representaciones de los Padres.*" FLACSO, Argentina. Unpublished manuscript (Masters' Thesis).
- Camargo, B., R. Camelo, S. Firpo, and V. Ponczek. 2011. "Test Score Disclosure and School Performance." Sao Paulo School of Economics Working Paper, Center for Applied Economics, Sao Paulo, Brazil.
- Carillo, P., M. Onofa and J. Ponce. 2010. "Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador." IDB Working Paper No. 223, Inter-American Development Bank, Washington, DC.
- Carneiro, P., and J. Heckman. 2004. "Human Capital Policy" in J. Heckman and A. Krueger, eds., *Inequality in America: What Role for Human Capital Policy?* MIT Press.
- Checchi, D., and H.G. van de Werfhorst. 2014. "Educational Policies and Income Inequality." IZA DP No. 8222
- Cerdan-Infantes, P., and C. Vermeersch. 2007. "More Time Is Better: An Evaluation of the Full-Time School Program in Uruguay." World Bank Policy Research Working Paper 4167.
- Coleman, J. 1966. "Equality of Educational Opportunity Study" ICPSR06389-v3. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.
- Cutler, D., and A. Lleras-Muney. 2008. "Education and Health: Evaluating Theories and Evidence." In J. House, R. Schoeni, G. Kaplan, and H. Pollack (eds.) *Making Americans Healthier: Social and Economic Policy as Health Policy*. New York: Russell Sage Foundation.

DeMartino, G., D.N. McCloskey, and G. Gamallo. 2011. "Mercantilización del Bienestar. Hogares Pobres y Escuelas Privadas." *Revista de Instituciones, Ideas y Mercados* N° 55, Octubre 2011, pp. 189-233, ISSN 1852-5970.

De Hoyos, R., J.M. Espino, and V. García. 2012. "Determinantes del Logro Escolar en México. Primeros Resultados Utilizando la Prueba ENLACE Media Superior," *El Trimestre Económico, Fondo de Cultura Económica*, vol. 0(316), pages 783-811, octubre-d.

DINIECE. 2015. *Anuario Estadístico Educativo. Relevamientos anuales 2007-2014. Ministerio de Educación de la Nación. Buenos Aires.*

DINIECE. 2011. *Relevamiento Anual 2011. Ministerio de Educación de la Nación. Buenos Aires.*

DINIECE. 2004. "PISA - Informe Nacional República Argentina." *OECD.*

Duflo, E., P. Dupas, and M. Kremer. 2012. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." NBER Working Paper No. 17939. National Bureau of Economic Research.

Duflo, E., R. Hanna, and S.P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review*, 102(4), 1241–1278.

Evans, D., and A. Popova. 2015. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." World Bank Policy Research Working Paper Series. Washington, DC.

Evans, M., J. Kelley, J. Sikora, and D. Treiman. 2010. "Family Scholarly Culture and Educational Success: Books and Schooling in 27 Nations." *Research in Social Stratification and Mobility*. Elsevier.

Fresoli, D., V. Herrero, R. Giulioldoli, and H. Gertel. 2007. "Incidencia de la gestión sobre el rendimiento escolar en la escuela argentina. El mensaje de las pruebas internacionales y nacionales." In *Anales de la Asociación Argentina de Economía Política*.

Fertig, M., and C.M. Schmidt. 2002. "The Role of Background Factors for Reading Literacy: Straight National Scores in the PISA 2000 Study." IZA Discussion Paper No. 545, Bonn: IZA.

Formichella, M. M. 2011. "¿Se debe el mayor rendimiento de las escuelas de gestión privada en la Argentina al tipo de administración?" *Revista de la CEPAL*, 105 (december), 151-166.

Fuchs, T., and L. Wößmann. 2004. "What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data." CESifo Working Paper Series 1235, CESifo Group Munich.

Fuller, B. 1987. "What School Factors Raise Achievement in the Third World?" *Review of Educational Research*, American Educational Research Association, Vol. 57, No. 3 (Autumn, 1987), pp. 255-292.

Gertler, P.J., H. Patrinos, and M. Rubio-Codina. 2012. "Empowering Parents to Improve Education: Evidence from Rural Mexico." *Journal of Development Economics*, Elsevier, vol. 99(1), pages 68-79.

Githau, B.N., and R.A. Nyabwa. 2008. "Effects of Advance Organiser Strategy during Instruction on Secondary School Students' Mathematics Achievement in Kenya's Nakuru District." *International Journal of Science and Mathematics Education*, 6, 439-457.

Glewwe, P., and H. Jacoby. 1994. "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana." *Journal of Human Resources*, University of Wisconsin Press, vol. 29(3), pages 843-864.

Glewwe, P.W., E.A. Hanushek, S.D. Humpage, and R. Ravina. 2014. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010." In *Education Policy in Developing Countries*, ed. Glewwe, P. University of Chicago Press: Chicago and London.

Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics*, 74, 251-268.

Glick, P., J.C. Randrianarisoa, and D. Sahn. 2011. "Family Background, School Characteristics, and Children's Cognitive Achievement in Madagascar," *Education Economics*, Taylor & Francis Journals, vol. 19(4), pages 363-396, February.

Gómez Schettini, M. 2007. "La Elección de Los No Elegidos: Los Sectores de Bajos Ingresos ante la Elección de Escuela en la Zona Sur de la Ciudad de Buenos Aires." In M. Narodowski and M. Gómez Schettini (eds.) *Escuelas y Familias. Problemas de Diversidad Cultural y Justicia Social*. Buenos Aires Prometeo Libros.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales. 2008. "Culture, Gender, and Math." *SCIENCE-NEW YORK THEN WASHINGTON-*, 320(5880), 1164.

Hanushek, E.A. 2003. "The Failure of Input-Based Policies." *The Economic Journal*, 113 (February), F64-F98.

Hanushek, E.A., S. Link, and L. Wößmann. 2013. "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA." *Journal of Development Economics*, 104, September 2013, pp. 212-232.

Hanushek, E.A., and L. Wößmann. 2007. "The Role of School Improvement in Economic Development." NBER Working Paper No. 12832, National Bureau of Economic Research.

Hanushek, E.A., and L. Wößmann. 2012. "Schooling, Educational Achievement, and the Latin American Growth Puzzle." *Journal of Development Economics* 99 (2): 497–512.

Hanushek, E.A., and S.G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100 (2): 267–71.

Heyneman, S., and W. Loxley. 1983. "The Effects of Primary School Quality on Academic Achievement Across Twenty-nine High and Low Income Countries." *American Journal of Sociology* 88, (1983): 1162-1194.

Kiboss, J.K. 2012. "Effect of Special E-Learning Program on Hearing-Impaired Learners' Achievement and Perception of Basic Geometry in Lower Primary Mathematics." *Journal of Educational Computing Research*, 46(1), 31-59.

Kremer, M., C. Brannen, and R. Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science*, 340 (6130), 297-300.

Kremer, M., E. Miguel and R. Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics*, 91(3) 437-456.

Krishnaratne, S., H. White, and E. Carpenter. "Quality Education for All Children? What Works in Education in Developing Countries." *3ie Working Paper 20*, International Initiative for Impact Evaluation.

Krüger, N. 2011. *La segmentación educativa en Argentina: Exploración empírica en base a PISA 2009, Actas de las XX Jornadas de la Asociación de Economía de la Educación.*

Lai, F., L. Zhang, Q. Qu., X. Hu, Y. Shi, M. Boswell, and S. Rozelle. 2012. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai." *Rural Education Action Project*, Working Paper No. 237, Stanford, CA.

Lee, J.-W., and R.J. Barro. 1997. "Schooling Quality in a Cross Section of Countries." *Economica* 38, no. 272: 465-88.

Linden, L.L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." J-PAL Working Paper, Abdul Latif Jameel Poverty Action Lab, Cambridge, MA.

Llach, J.J., C. Adroque, and M.E. Gigaglia. 2009. "Do Longer School Days have Enduring Educational, Occupational or Income Effects? A Natural Experiment on the Effects of Lengthening Primary School Days in Buenos Aires, Argentina." *Economía* 10, no. 1 (2009): 1-39.

- Louw, J., J. Muller, and C. Tredoux. 2008. "Time-on-Task, Technology and Mathematics Achievement." *Evaluation and Program Planning*, 31 (1) (Feb): 41-50.
- Malamud, O., and C. Pop-Eleches. 2011. "Home Computer Use and the Development of Human Capital." *Quarterly Journal of Economics*, 126, 987-1027.
- Majoribanks, K. 1994. "Perceptions of Parents' Involvement in Learning and Adolescents' Aspirations." *Psychological Reports: Volume 75*, Issue, pp. 192-194.
- Marchionni, M., F. Pinto, and E. Vazquez, 2013. "Determinants of the Inequality in PISA Test Scores in Argentina." MPRA Paper 56421, University Library of Munich, Germany.
- Marshall, J.H., and A.M. Sorto. 2012. "The Effects of Teacher Mathematics Knowledge and Pedagogy on Student Achievement in Rural Guatemala." *International Review of Education*, 58(2), 173-197.
- McEwan, P. 2012. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research* 20 (10): 1-42.
- McDonald, K. 2008, revised 2014. "PV: Stata Module to Perform Estimation with Plausible Values." Statistical Software Components from Boston College Department of Economics .
- Mizala, A., P. Romaguera, and M. Urquiola. 2007. "Socioeconomic Status or Noise? Tradeoffs in the Generation of School Quality Information." *Journal of Development Economics*, Elsevier, vol. 84(1), pages 61-75, September.
- Mizala, A., and M. Urquiola. 2013. "School Markets: The Impact of Information Approximating Schools' Effectiveness." *Journal of Development Economics*, 103, 313-335.
- Mo, D., J. Swinnen, L. Zhang, H. Yi, Q. Qu, M. Boswell, and S. Rozelle. 2012. "Can One Laptop per Child Reduce the Digital Divide and Educational Gap? Evidence from a Randomized Experiment in Migrant Schools in Beijing." Rural Education Action Project Working Paper 233, Stanford, CA.
- Mo, D., L. Zhang, R. Lui, Q. Qu, W. Huang, J. Wang, Y. Qiao, M. Boswell, and S. Rozelle. 2013. "Integrating Computer-assisted Learning into a Regular Curriculum: Evidence from a Randomized Experiment in Rural Schools in Shaanxi." Rural Education Action Project Working Paper 248, Stanford, CA.
- Muralidharan, K., and V. Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *The Journal of Political Economy*, 119, 39-77.

- Muralidharan, K., and V. Sundararaman. 2010. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *The Economic Journal*, 120, no. 546: F187-F203.
- Murnane, R.J., and A.J. Ganimian. 2014. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations." NBER Working Paper No. 20284, National Bureau of Economic Research, Cambridge, MA.
- Narodowski, M., and M. Moschetti. 2015. "The Growth of Private Education in Argentina: Evidence and Explanations." *Compare: A Journal of Comparative and International Education*, 45:1, 47-69, DOI: 10.1080/03057925.2013.829348
- Narodowski, M. 2002. "Monopolio Estatal y Elección de Escuela en la Argentina." In M. Narodowski et al, *Nuevas Tendencias en Políticas Educativas. Estado, Mercado y Escuela*. Buenos Aires.
- OECD. 2013). "PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices" Volume IV, PISA, OECD Publishing.
- OECD. 2014. "PISA 2012 Results: What Students Know and Can Do—Student Performance in Mathematics, Reading and Science." Volume 1. Paris: OECD Publishing. doi :10.1787/19963777.
- Pelletier, R., and A.H. Normore. 2007. "The Predictive Power of Homework Assignments on Student Achievement in Mathematics." In S.M. Nielsen and M.S. Plakhotnik (eds.), *Proceedings of the Sixth Annual College of Education Research Conference: Urban and International Education Section* (pp. 84-89). Miami: Florida International University.
- Petrosino, A., C. Morgan, T.A. Fronius, E.E. Tanner-Smith, and R.F. Boruch. 2012. "Interventions in Developing Nations for Improving Primary and Secondary School Enrollment of Children: A Systematic Review." *Campbell Systematic Reviews* 2012:19.
- Ravallion, M., and S. Chen. 2003. "Measuring Pro-Poor Growth." *Economics Letters*, 78(1), 93-99.
- Santibañez, L., R. Abreu-Lastra, and J.L. O'Donoghue. 2014. "School-Based Management Effects: Resources or Governance Change? Evidence from Mexico." *Economics of Education Review*, 39, pp. 97-109.
- Santos, M. 2007. "Quality of Education in Argentina: Determinants and Distribution using PISA 2000 Test Scores." *Well-Being and Social Policy*, vol. 3, No. 1, Mexico City, Inter-American Conference on Social Security/Ibero-American University.
- SEDLAC. 2015. Socio-Economic Database for Latin America and the Caribbean. Accessed June 2015.

Todd, P.E., and K.I. Wolpin. 2003. "On The Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal*, Royal Economic Society, vol. 113(485), pages F3-F33, February.

Tosoni, M., and A. Natel. 2010. "Elegí esta escuela primaria para mi hijo." *Las elecciones educativas de las familias desectores populares. VI Jornadas de Sociología de la UNLP. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Departamento de Sociología, La Plata.*

UNESCO, 2015. UNESCO Institute for Statistics—Statistics. UNESCO, Paris. <http://www.uis.unesco.org/Datacentre/Pages/instructions.aspx?SPSLanguage=EN> (accessed 12 November 2015).

World Bank. 2011. "Education Sector Strategy 2020: Learning for All." World Bank Group: Washington, DC.

World Bank. 2014. "Argentina—Country partnership strategy for the period of FY2015-18. World Bank Group: Washington, DC.

Yayan, B., and G. Berberoglu. 2004. "A Re-Analysis of the TIMSS 1999 Mathematics Assessment Data of the Turkish Students." *Studies in Educational Evaluation*, 30, 87-104.

Ziliak, S.T., and D.N. McCloskey. 2014. "Lady Justice v. Cult of Statistical Significance: Oomph-less Science and the New Rule of Law." In the forthcoming *Oxford Handbook on Professional Economic Ethics*, ed.

Annex A – Statistical Analysis Using PISA

The Program for International Student Assessment (PISA) was launched by the OECD with the objective of providing a comprehensive set of data on students’ performance and associated background, and schools and institutions characteristics comparable across countries. The first edition of the survey was conducted in 2000, with 32 countries participating. Eleven non-OECD countries, including Argentina, joined in 2001 and 2002.

The PISA survey design is very complex, and its specific technical features and potential limitations for statistical analysis are usually overlooked in the literature. Getting a clear understanding of the survey design is critical to undertaking an accurate statistical analysis and correct interpretation of results. This annex briefly presents some survey characteristics that have to be taken into account when using PISA data for statistical analysis.

PISA is an age-based survey, assessing 15-year-old students in grade 7 or higher. It provides data on their achievement in reading literacy, mathematics literacy, and science literacy. Each PISA round assesses one of the three subject areas in depth, with the other two assessed as minor areas (Table A 1). The test takes a literacy perspective and focuses on the extent to which students can apply the knowledge and skills they have learned at school when confronted with real-world situations. This “yield” approach, focusing on the application of competencies, distinguishes PISA from other international learning assessments (for instance, TIMMS/PIRLS), which aim at measuring school-based curricular attainment more closely.

Table A 1: PISA Survey by Main Subject Area Assessed

<i>Edition</i>	<i>Main subject</i>
2000	Reading
2003	Mathematics
2006	Science
2009	Reading
2012	Mathematics

Source: PISA-OECD

PISA uses a two-stage sampling. In the first stage, schools are drawn from the population of schools with a probability proportional to their size, and in the second stage a fixed number of children in the schools are sampled. This sampling procedure implies that students are not identified by a pure random selection: on the contrary, they are somewhat “clustered” by school. Standard errors that are not clustered at the school level would then underestimate the true population variation. Therefore, the population weight for a student must reflect the probability of his/her school being selected, as well as the probability that he/she was selected within the school.

To correct for this, PISA adopts the Balanced Repeated Replication (BRR) variant of the jackknife technique, in its modification developed by Fay (1989).¹⁷ In PISA, the BRR approach considers 80 replicate samples, with the resulting 80 weights included in the raw dataset. When

¹⁷For more details on these techniques to calculate standard errors in complex designed sample, see McCarthy (1969), Fay (1989), and Judkins (1990). The Fay constant assumed by PISA equals 0.05.

conducting data analysis, these weights have to be used to take into account the PISA two-stage sample design.

Final test scores are expressed in five plausible values (PVs). Instead of obtaining a point estimate for student's ability, PISA reports a range of possible values for it, with an associated probability for each of those values estimated. This approach aims at correcting potential measurements errors, as well as potential bias arising when translating a continuous variable (i.e. cognitive skills) in a discontinuous one (i.e. number of correct answers). As described by Wu and Adams (2002), "the simplest way to describe plausible values is to say that they are a representation of the range of abilities that a student might reasonably have" given his or her answers to the different test items. Final mean test score is the simple average of the five plausible values.

The PV approach implies that one cannot compute variance of test scores as the simple dispersion of the final mean test score, as would be the case for a test reporting one score per student. As explained in detail in the PISA Technical Report (2010) and in the PISA SPPS Full Manual (2010), this would lead to an underestimation of the overall variance, increasing the probability of Type I error. Rather, the BRR method has to be applied to all five PVs, to compute the sampling variance of each of them. The average of the five sampling variances will then give us the final sampling variance. Furthermore, the use of PV introduces a new source of uncertainty, since the mean of each PV differs from the final estimate of total mean score. This additional measurement error, or imputation variance, has to be taken into account, otherwise the final sampling variance would underestimate the true variance. Hence, the final variance has to reflect both final sampling variance and imputation variance. The following steps summarize the main steps to calculate final test score population variance.

1. Estimate the mean and sampling variance for each of the five PVs of PISA, by using final student weight (w_{fstuwt}) and the 80 BRR weights (w_{fstr1} - w_{fstr80}). Note that this means that 81 estimates for each PV are needed in order to get the final estimate and its standard error.
2. The overall mean estimate is the simple average of the five mean estimates for the same number of PVs:

$$\hat{\mu} = \frac{1}{5} (\hat{\mu}_{pv1} + \hat{\mu}_{pv2} + \hat{\mu}_{pv3} + \hat{\mu}_{pv4} + \hat{\mu}_{pv5})$$

3. Sampling variance is the simple average of the five sampling variances across the five PVs:

$$\hat{\sigma}_{sampling} = \frac{1}{5} (\hat{\sigma}_{pv1} + \hat{\sigma}_{pv2} + \hat{\sigma}_{pv3} + \hat{\sigma}_{pv4} + \hat{\sigma}_{pv5})$$

4. Estimate the imputation/measurement error variance:

$$\hat{\sigma}_{Imp} = \frac{1}{M-1} (\hat{\mu}_i - \hat{\mu})^2$$

where M is the number of plausible values.

5. Get the final variance for test scores:

$$\hat{\sigma}_{final} = \hat{\sigma}_{sampling} + \left(1 + \frac{1}{M}\right) \hat{\sigma}_{Imp}$$

Note that there are specific statistical software commands that can help automate these steps. In Stata, options range from the “*pv*” command written by McDonald (2003, 2013) and the “*pvpisa*” by Lauzon (2004). Some authors opt for a more common “*svy brr*” option (see Kreuter and Valliant, 2007, and Santos, 2003), disregarding imputation variance. Depending on the type of scope of the analysis, however, the imputation variance could turn out to be relevant and potentially modify results.

PISA’s design follows Item Response Theory (IRT) techniques. Results are then calibrated at national and, in a following step, international levels. In 2000 the reading reporting scale was adjusted so that the mean and standard deviation of the PISA 2000 scores was 500 and 100 respectively, for the equally weighted 27 OECD countries that participated in PISA 2000 and had acceptable response rates (Wu and Adams, 2002). This reading scale was maintained in the subsequent editions to ensure comparability of test scores over time. On the other hand, scales for mathematics and science were defined *ex novo* in 2003 and 2006, respectively. For this reason, only reading scores are strictly comparable over the period 2000-2012.

Clearly, adopting the same reporting scale is not sufficient in itself to ensure comparability. To ensure that tests from different editions can be scaled on the same continuum, or, in other words, to replicate the same level of test difficulty over time, some common “linking” or “anchor” items are included in the different surveys. The choice of which items to use as linking, as well as their number, introduces an additional source of variation. This means that if an alternative set of linking items had been chosen, the results would be slightly different. The uncertainty that results from the link-item sampling is referred to as linking error, which must be taken into account when making comparisons between the results from different PISA editions.

As the PISA survey items are clustered in units, the computation of the linking error estimates is quite complex. The PISA SPSS Full Manual (2010) provides a list of the linking error estimations for each two test surveys, in each subject. The PISA 2012 Results (2013) provides an update to link 2012 results to previous editions. Linking error estimations are summarized in

Table A 2. To test, say, the statistical significance in a change in scores between t and $t+1$, one will have to estimate the standard error according to:

$$SE = \sqrt{\sigma_t^2 + \sigma_{t+1}^2 + \sigma_{Linking}^2}$$

Failing to account for linking error would lead to a considerable underestimation of variance, with an increase in probability of estimating significant changes in results when, instead, changes may not be significant (Type I error).

Table A 2: Linking Error for Comparisons of Performance between PISA Editions

	Link error on PISA scale
Reading – 2000-2003	5.307
Reading – 2000-2006	4.976
Reading – 2000-2009	4.936
Reading - 2000-2012	5.923
Reading – 2003-2006	4.474
Reading – 2003-2009	4.088
Reading – 2003-2012	5.604
Reading – 2006-2009	4.069
Reading – 2006-2012	5.580
Mathematics – 2003-2006	1.990
Mathematics – 2003-2009	1.333
Mathematics – 2003-2012	1.931
Mathematics – 2006-2009	1.382
Mathematics – 2006-2012	2.054
Science – 2006-2009	2.566
Science – 2006-2012	3.512
Science – 2009-2012	2.006

Source: PISA - OECD

Annex B – Education Production Function without Institutional Controls

Independent Variables	2000	2006	2009	2012
<i>Student's characteristics</i>				
Gender (female)	25.625***	40.415***	24.498***	25.293***
Not enrolled in 10th grade (over age)	-82.981***	-78.963***	-66.160***	-59.178***
<i>Socioeconomic background</i>				
Wealth index	1.35	15.903***	10.118***	1.57
Mother's highest level of education				
<i>Complete primary school</i>	-23.917**	-9.861	4.891	17.433**
<i>Complete upper secondary school</i>	-15.694	1.886	19.497**	16.987**
<i>Complete lower secondary school</i>	-2.446	2.945	17.129***	28.773***
<i>Complete tertiary</i>	-4.713	6.156	31.195***	42.465***
Father's highest level of education				
<i>Complete primary school</i>	-12.634	7.294	5.344	7.208
<i>Complete upper secondary school</i>	-12.833	17.017	11.073	9.263
<i>Complete lower secondary school</i>	2.148	6.085	6.866	8.781
<i>Complete tertiary</i>	2.475	20.552**	9.774	6.474
Books at home				
<i>11-100</i>	15.849*	18.676***	14.551***	20.303***
<i>more than 100</i>	40.230***	37.067***	35.463***	37.220***
<i>School characteristics</i>				
Public school	-26.648**	-42.943***	-24.632***	-45.691***
Rural area	-25.892***	-5.885	-26.644**	7.38
School size				
<i>200-500 students</i>	5.899	46.376***	28.824**	-1.559
<i>500-1,000 students</i>	18.146	46.942**	37.840***	4.312
<i>> 1,000 students</i>	17.377	59.044***	34.988**	4.34
Student-teacher ratio (full time + part-time)	0.324	-0.321	-0.064	0.072
Proportion of teachers with tertiary education, %	.879*	0.096	.425*	.362***
Computer with web connection, per student (x 100)	1.404	1.234	3.554***	.539*
Shortage/Inadequacy of instructional material	-8.698	-19.127	-1.575	-4.003
Constant	442.284***	373.518***	369.807***	383.565***
N	2,459	2,834	2,865	3,373
R2	0.36	0.37	0.42	0.36

Notes: All estimation results are accounting for PISA's survey design as explained in Annex A. As usual, * p<.10, ** p<.05, *** p<.01. The references category for mother's and father's education level is unfinished primary; the reference category for number of books at home is less than 11; and the reference category for school size is fewer than 200 students.

Annex C – Education Production with Institutional Controls Defined by Hanushek, Link, and Wößmann (2013)

Independent Variables	2000	2006	2009	2012
<i>Student's characteristics</i>				
Gender (female)	24.844***	38.391***	23.964***	25.410***
Not enrolled in 10th grade (over age)	-73.476***	-75.148***	-66.000***	-58.789***
<i>Socioeconomic background</i>				
Wealth index	-1.996	15.405***	9.921***	1.547
Mother's highest level of education				
<i>Complete primary school</i>	-24.557*	-11.802	5.185	16.888**
<i>Complete upper secondary school</i>	-17.389	0.002	19.949**	16.371**
<i>Complete lower secondary school</i>	0.379	0.203	17.858***	28.365***
<i>Complete tertiary</i>	-4.271	3.83	31.475***	41.824***
Father's highest level of education				
<i>Complete primary school</i>	-19.059	7.335	6.008	7.06
<i>Complete upper secondary school</i>	-22.941	17.279	11.738*	9.361
<i>Complete lower secondary school</i>	-6.123	6.973	7.155	8.865
<i>Complete tertiary</i>	-11.82	21.659**	10.05	6.506
Books at home				
<i>11-100</i>	17.913***	18.002***	14.748***	20.487***
<i>more than 100</i>	37.480***	37.999***	34.316***	37.311***
<i>School characteristics</i>				
Public school	15.618	-67.471**	-12.643	-48.280***
Rural area	-14.349	-11.967	-26.215**	8.94
School size				
<i>200-500 students</i>	-10.891	45.137***	29.847**	1.336
<i>500-1000 students</i>	9.573	44.616**	40.475***	7.886
<i>> 1000 students</i>	1.886	54.379**	37.210**	8.886
Student-teacher ratio (full time + part-time)	0.384	-0.485	-.322**	-0.208
Proportion of teachers with tertiary education	0.77	0.106	.390*	.361***
Computer with web connection available per student	2.113	1.192	3.381***	.538*
Shortage/Inadequacy of instructional	-22.224***	-19.896	0.498	-4.815
<i>Institutional setting</i>				
Autonomy in deciding on teachers' salaries	-19.216	-10.348	12.322	-5.239
Autonomy in budget and personnel decisions	47.762***	-32.366	25.727	-1.695
Autonomy in deciding on academic content	22.101*	10.641	8.857	9.507
Autonomy in monitoring students' performance	-18.585	-23.757	-0.201	-0.467
Constant	410.590***	418.354***	348.346***	384.475***
N	1,892	2,834	2,865	3,373
R2	0.37	0.38	0.43	0.37