“Scaling Up Preschool Education: A Semiparametric Estimation of its Effect on Early Learning Outcomes in Peru”

Master Thesis presented to the Department of Economics at the
Rheinische Friedrich-Wilhelms-Universität Bonn

In Partial Fulfillment of the Requirements for the Degree of Master of Science (M.Sc.)

Supervisor:
Prof. Dr. LENA JANYS

Submitted in October, 2021 by:

NICOLAS EDUARDO FAJARDO ACOSTA
Matriculation Number
3299422

Contents

Abstract

In this document, I present an analysis of the expansion of the national preschool education in Peru, trying to exploit the exogenous variation generated by the policy after the 2004's Education Reform. Consequently, using a family-fixed effects design that allows heterogeneity, I seek to identify the effect through the variation in siblings' exposure to the rollout of preschools across Peru. In theory, such an empirical strategy allows me to overcome selection concerns, as well as to account for external validity issues. Unfortunately, I find that the estimation framework, although appropriate in its theoretical concerns, cannot be used in this particular setting due to an important lack of power. I analyze the possible explanations and suggest new avenues to try to recover the Average Treatment Effect [ATE] for this particular example.

KEYWORDS | *Preschool, Policy Evaluation, Family Fixed Effects, Propensity Score Matching, Education in Developing Countries, Quasi-experiment*

CHAPTER 1
Introduction

During the last decades, Early Childhood Development [EDC] got the attention of global development policies and programs in part because developments in neuroscience and developmental biology shed light upon the importance of it, explaining that early opportunities, as well as early adversities, have crucial roles determining later educational, social, and economic outcomes of the individuals, but also at the societal level (Duncan & Magnuson, 2013). From an economic perspective, as emphasized by Nores (2020), given their broader social benefits, early childhood programs are seen as public goods worth investing in by governments and not dependent solely on individual family choices or constraints. This led economics to approach EDC from three different angles; (i) within market failures, with positive social externalities but constrained by credit and financing opportunities. (ii) as a way to reduce income inequalities (iii) as a human capital accumulation promoter, where the objective is to optimize interventions during children's life cycle and achieve higher latter productivity, and long-lasting benefits in health, crime, and other areas.

According to Nores and Barnett (2010), in the short and medium-term, early childhood interventions, particularly those that focus on the developmental and cognitive aspects of childhood, have positive impacts on the cognitive development of children, their school progress, and educational attainment. Its proven effectiveness and its far-reaching consequences have been considerable attractors to implement large-scale national programs, especially in low- and middle-income countries (Britto, Yoshikawa, et al., 2013). However, and despite the efforts, the provision does not equate quality, a fact that is of central concern, moreover if the goal is reducing disparities and fostering human capital formation (Barnett, 2011). For developed countries, and particularly the United States, large-scale preschool programs have had strong positive effects across several short and run-long outcomes (Yoshikawa et al., 2013). Unfortunately, much less is known about the effectiveness of national preschool programs to influence outcomes and narrow socioeconomic gaps in developing countries (Richter et al., 2017). Therefore, it is crucial to evaluate the effectiveness of such programs, not only as a cost-benefit analysis, but also —in the words of (Majerowicz Nieto, 2019)— to seek "understanding of which components of preschool matter the most, and therefore [realize] how national preschool programs should be structured".

In this document, I present an analysis of the expansion of a national preschool program in Peru, applying a framework that seeks to exploit the exogenous variation generated by the expansion of the preschool system after the 2004 Education Reform. The expansion "sought to close the achievement gaps between students and improve overall learning outcomes by better preparing students for primary school" (Majerowicz Nieto, 2019). Consequently, using a family-fixed effects design that allows heterogeneity, I try to identify the effect through the variation in siblings' exposure to the rollout of preschools across Peru. Such an empirical strategy allows me to overcome selection concerns when comparing across towns. Namely, that assigned towns could differ systematically from non-assigned ones given political connections, accessibility conditions, or, more broadly, socioeconomic and cultural factors. Assuming that the slow establishment of preschools across towns generated exogenous variation within families conditional on other covariates, I can compare the outcomes of the younger and older siblings, where the younger had access to a new preschool but the older was not eligible to benefit from the program. Unfortunately, I find that the estimation framework, although appropriate in its theoretical concerns, cannot be used in this particular setting, suggesting new avenues to try to recover the Average Treatment Effect [ATE].

The analysis continues as following; on what is left in this chapter, I revisit the concept of Early Child Development, show its current state in the world, and detail its implications for public policy in developing countries. Later, I give a short historical and descriptive overview of the Peruvian education sector and its associated policies. In chapter 2, I analyze the data sources and relate, for each relevant variable, the motivation, challenges, and procedures to gather, clean, and use the available data. Next, in chapter 3, I explain how I attempted to identify the effect, addressing both internal and external validity concerns, and argue why the current application theoretically fulfills both identification and statistical assumptions for the credible recovery of the ATE (ITT). Finally, in chapter 4, I discuss that the empirical reality is different and give some possible explanations about the lack of power of the method for this particular analysis.

Early Childhood Development

Across all disciplines involved, it has been heavily emphasized, that early life experiences and contexts are crucial to shape successful later-in-life outcomes. This linkage is key, not only as a way to improve the individual and familiar well-being, but also because it produces a myriad of positive social welfare spillovers, such as school success and completion, higher earnings and productivity, active participation in communities and society, and reduced odds of delinquency, crime, and chronic and non-communicable diseases. Those complementary chain effects, place EDC at the center of the public policy debate as a outstanding area to reduce inequality and achieve stable paths of economic and social development, inasmuch as —evident as it sounds— children are the adults of the future. EDC encompasses all ecological characteristics —health, nutrition, protection, care and education— that configure a child's life outcomes on later life stages. More specifically, according to Britto, Engle, et al. (2013), the ECD comprises two complementary scopes: the age and the domains of development of a human being.

In terms of age, as shown by Britto, Engle, et al. (2013), the Early Childhood generally covers the time span between conception until 8 years of age, or until the transition to school is completed. Within that interval, three main phases can be identified: (i) The first three years, when nutrition, health, and stimulation play a forefront role. (ii) Between three and five years, when in addition, children start to interact with a broader community, including peers and other caretakers, and appear to be more vulnerable to violence, abuse, and neglect within their homes. (iii) The End of Early Childhood, when group learning and socialization opportunities become central to the developmental process, reflecting a fundamental shift in children's learning and interaction with the environment, and setting a lifelong foundation that determines the acquisition, sustaining, and promotion of their own "human capital" all over the life.

In terms of domains of development, Britto, Engle, et al. (2013) highlight that the conceptualization of the EDC includes, but it is not limited to, physical health and motor development, cognitive and language skills, social and emotional functioning, ethical and spiritual development, and sense of national or group identity. Although there are cases of plasticity, this multidimensional approach is specially important, since it is also assumed that the joint results of health, nutrition, early stimulation, social and emotional interactions with caregivers and peers, play and learning opportunities, and protection from violence and neglect are irreversible, in other words, the outcomes on later-in-life stages are path dependent with respect to the overall undergone experiences before nine years of age.

However, even when the evidence is strong, the overall international situation for the majority of children and their families is not good. As reported by UNICEF (2019), "At least one in three children under the age of five is stunted, wasted or overweight and, in some cases, suffers from a combination of two of these forms of malnutrition.", while according to Richter et al. (2020), in 2020, one out of six children does not receive the benefits of early stimulation and responsive care by adults at home; and about every two out of five children aged 1–4 years experience violent discipline by their caregivers. This panorama worsens quickly if we disaggregate the data into levels of income. Of all the world's children, only about 8% are born in a high-income country, while the other 92% of children live in low- and middle-income countries (Engle et al., 2013). Lu et al. (2020) show that in 2018:

> On average, 63.2% of children in 85 countries were neither exposed to stunting nor to extreme poverty, with the highest proportion found in Europe and Central Asia (88.3%) and the lowest in sub-Saharan Africa (45.2%). Across 65 countries with available data, fewer than two fifths (38.9%) of children ever attended an [Early Childhood Education (ECED)] programme, with the highest level in East Asia and the Pacific (67.4%) and lowest in sub-Saharan Africa (24.1%). Just over two thirds (69.1%) of children were exposed to home stimulation defined as adequate, with the highest percentage in Europe and Central Asia (90.1%) and lowest in sub-Saharan Africa (46.9%). In the 60 countries with [Early Childhood Development Index (ECDI)] data, 75.1% of children were rated as developmentally on track according to the ECDI cut-offs, with the lowest proportion in sub-Saharan Africa (60.7%) (p. 4).

Lu et al. (2020) also find important within countries variations when considering the rural-urban gap (favoring always the urban children) and household wealth gap (favoring always the children in the richest quintiles). Gender differences were in general negligible, with the notable exception of girls being less exposed "(...) to stunting or extreme poverty than boys in Latin America and the Caribbean (−1.6 pp with 95% CI −2.2 pp and −1.0 pp), sub-Saharan Africa (−2.5 pp with 95% CI −3.2 pp and −1.9 pp), and in higher income groups relative to low-income groups", while having higher scores on the ECDI "in the Middle East and North Africa (6.2 pp with 95% CI 10.4 pp and 2.0 pp), South Asia (4.6 pp with 95% CI 8.5 pp and 0.7 pp) and sub-Saharan Africa (4.7 pp with 95% CI 8.5 pp and 0.9 pp)".

Public policy and Early Childhood in Developing Countries

Given the importance of EDC, but also because of the marked differences in the general standing of children between countries, in the last 40 years, EDC has played a central role in national and international public policy contexts, being part of the Sustainable Development Agenda 2030 (Britto, Engle, et al., 2013). Consequently, after years of debate, analysis, and implementation, and considering the diverse factors involved in EDC, a wide range of public policy approaches has been developed to address the most important problems: while the health and nutrition sector focuses on survival, disease prevention, and growth, with maternal and child health usually highlighted for special attention; the education sector is interested in learning and primary school preparation, and the protection sector focuses especially on vulnerable populations

(Behrman & Urzúa, 2013). Focusing the analysis on the education sector, Nores (2020) argue that its main objective can be summarized as to "ensur[e] that all children have access to quality early childhood development, care and pre-primary education so as to be ready for primary education". However, in poor, unequal and vulnerable contexts, such a basic tenet constitutes a serious challenge to both research and practice on Early Childhood Education, considering the funding limitations, the widespread economic, political, and social risks, and a general lack of governance at local and regional levels (Britto, Yoshikawa, et al., 2013).

In that sense, it is not surprising that the provision of Early Childhood Education and Care differ significantly among countries, not only because each program supports different aspects of childhood, but because they take place in formal, informal, and non-formal settings, in arrangements ranging from center-based to formal preschool education, to parent/community-based arrangements or home-visiting (Barnett, 2011). Following Trawick-Smith (2019), such characteristics are empirically and theoretically important, especially in poorer contexts, as the programs are themselves embedded in their particular social, cultural, and economic contexts, and therefore, reducing disparities would not be only a matter of cost-benefit analysis, but the result of wider social and political changes. Nonetheless, most of the research and policy focus on the cognitive and non-cognitive effects of the interventions, seeking to justify the public spending supported by the estimated marginal gains on the outcomes, which in practice puts preschool, and primary and secondary education at the same public policy hierarchical level, being only differentiated by their "potential" effects and underlying costs (Nores, 2020). Consequently, to assess the short-run effects, the analyses often regard children's cognitive development, test performance, intelligence quotients, school progress, educational attainment, nutritional and socialization indicators, and even their mothers' labor supply. Instead, for the long run, direct and indirect effects like social and emotional skills, educational success, criminal activity, and integration rates are assessed. However, as Behrman and Urzúa (2013) discuss:

> Some systematic evidence suggests that ECD programs in developing countries may have substantial effects in some contexts, probably more so for children from poorer families. However, there are large gaps in the literature regarding the benefits of ECD and ECD programs in these countries and how the benefits are distributed among children from different types of families. There are even greater gaps in measuring real resource costs and in understanding alternative financing and management possibilities and related appropriate policies (p. 136).

Peruvian Education Sector in Context

Since the 1980s, Latin America has undergone important educational reforms at all education levels. Among them, as Gregosz et al. (2014) state, three historical phases can be identified. The first one, occurred during the 1980s, aimed to decentralize the management of the public education —giving provinces and administrative regions more autonomy— and to change the funding schemes for access —establishing credit and grant programs—. The second, starting after the trade liberalization that swept over Latin America in the late 1980s and early 1990s, aimed to improve the pedagogical processes in the classroom, striving to increase the levels of efficiency, quality and autonomy through significant investments on training, infrastructure and materials. Finally, in the 2000s onward, as governments started to deploy national-level standardized evaluation mechanisms, for both students and teachers, and strengthen the accountability of the

Table 1.1 Structure of the Peruvian School Education System.

LEVEL	PRESCHOOL		PRIMARY SCHOOL						SECONDARY SCHOOL				
STAGE	I	II	III		IV		V		VI		VII		
Grade	*Cuna**	*Jardín/Pronoei***	*1st*	***2nd***	*3rd*	***4th***	*5th*	*6th*	*1st*	***2nd***	*3rd*	*4th*	*5th*
Age	0–2	3–5	6	**7**	8	**9**	10	11	12	**13**	14	15	16

Age denotes the age cohort that is legally expected to attend to each corresponding grade.
The **boldfaced** items indicate when the National Standardized Tests [ECE] are administered.

* Corresponds to the Spanish name given to the childcare facilities, meaning "Crib". There are also community and family programs aimed to support the 1st stage called SET, PIET, PIETBAF and PAIGRUMA. However, since the focus of this paper lies in later stages, less emphasis is put on those.

** Designates each of the preschool programs. *Jardín* (Kindergarten) refers to the regular program, while *Pronoei* refers to the community program.

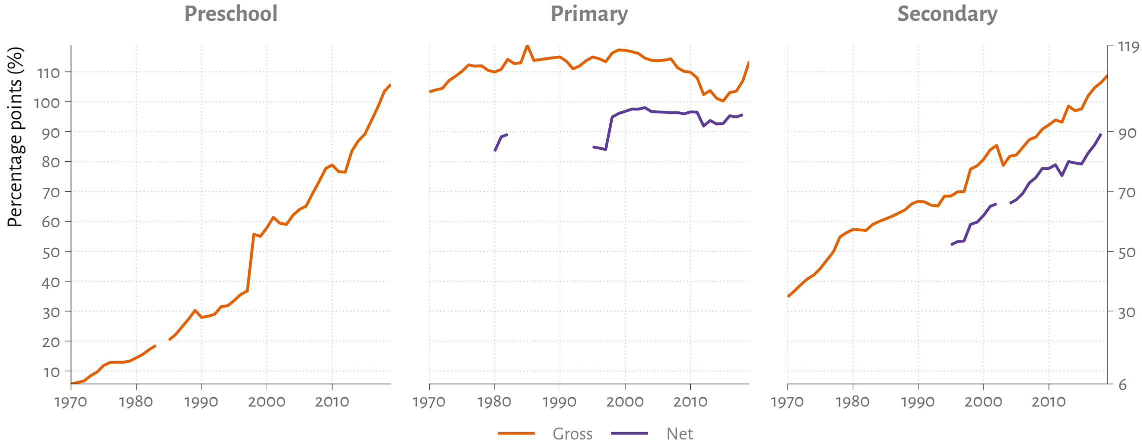
education institutions. Nevertheless, today, after decades of efforts, Latin America remains with serious access deficits —specially in secondary education—, low teaching quality, sizable infrastructure underinvestment, and very low rates of student learning compared to its international counterparts. Moreover, the regional, social and economic disparities still play an important role shaping the educational outcomes across the region: poorer, rural and ethnic students exhibit particularly lower learning rates. For example, 12% of economic disadvantaged students achieve at least a minimum level of reading comprehension, whereas 52% of their richer counterparts achieve the aforementioned.

Such disparities in the learning rates are remarkably important in a context where: (i) parental education is the main driver of the observed intergenerational persistence in terms of inequality of opportunities, (ii) and educational attainment explains an increasing trend of positive social mobility in one of the most unequal regions of the world (Neidhöfer et al., 2018; Torche, 2014). Therefore, the implied "double causality", where educational outcomes seem to be determined by socio-economic and demographic factors, and education simultaneously reduces the socio-economic intergenerational gaps, constitute —just like in the international experience— important idiosyncratic challenges for Latin American countries, their policy making and its specific implementations.

The panorama in Peru is not different from the regional stylized facts. During the last 20 years, Peru has experienced significant economic growth rates, which unfortunately have not been translated into improved human capital accumulation. For instance, Peru is still between the five worse overall performers in the Program for International Student Assessment [PISA] tests, while consistently holding the last places according to the yearly Global Competitiveness Reports of the World Economic Forum in terms of primary school and overall education quality—especially in fields like mathematics and science—. According to Ortiz Chávez and Pastor Vargas (2014), such an insufficiency can be explained by the general condition of the Peruvian education system, ranging from its organization to its coverage, available infrastructure, educational staff, and public investments. In that regard, Guadalupe et al. (2017) discuss that the basic structure of the Peruvian education system has not changed significantly over the last 110 years. The most important changes are linked to the expansion of compulsory education, which today includes preschool, primary and secondary education, policy which after 15 years changed the dynamics of enrollment growth in school education: from a system in constant expansion and, therefore, in need of increasing resources, it has become a system of stable size, with a certain tendency toward reduction, providing an interesting opportunity to restructure the provision of educational services in the light of its low quality according to international standards.

Nowadays, the school education system in Peru is subdivided in levels, stages, and

Figure 1.1 Enrollment Rates per School Level (1970–2019).



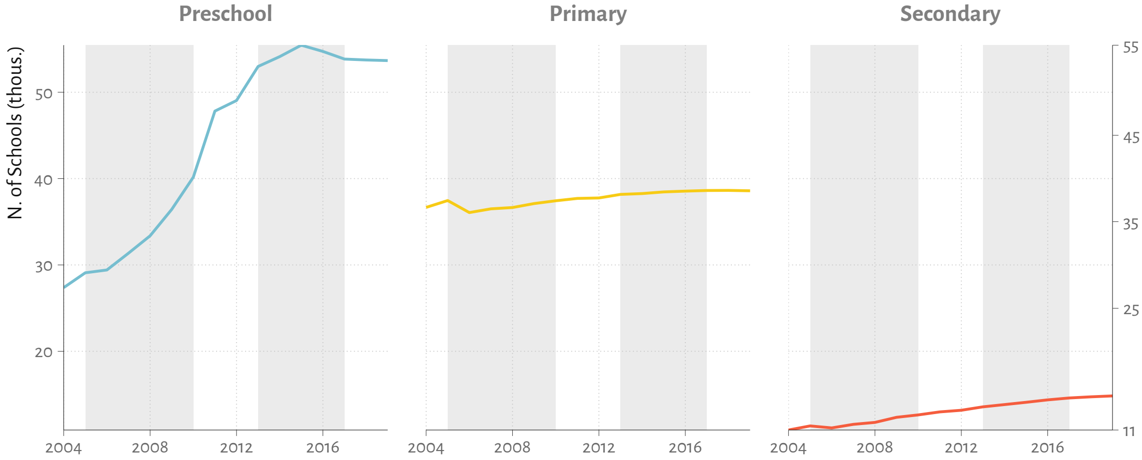
Source: UNESCO Institute for Statistics. Data as of September 2020. Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown (Therefore it can go over 100%). Net enrollment rate is the ratio of children of official school age who are enrolled in school to the population of the corresponding official school age. The difference between both measures reflects late enrollment, early enrollment, and retention. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

grades (see table 1.1). While the levels refer to the UNESCO's International Standard Classification of Education —merging Level 2 and 3 under "Secondary School"—, the stages categorize different learning outcomes depending on each age group and its associated cognitive and non-cognitive skills.[1] Furthermore, as seen in the fig. 1.1, the gross enrollment rates have risen steadily for both preschool and secondary education, breaking the 100% mark in 2018. Such a phenomenon should not be misunderstood, as gross enrollment rates include under- and overaged as well as retained students, which in turn could highlight inefficiencies within the education system. However, in general, during the first two decades of the XXI century, Peru has improved substantially its school education coverage at all levels, but more importantly, has kept its primary net enrollment rates around 95% and thus, effectively setting the primary education as an universal right.

Although the enrollment rates show uplifting trends, it does not automatically imply, that students attend to their respective schools. Thus, to gain deeper understanding, it is also important to contrast enrollment and attendance rates. In general, as Ortiz Chávez and Pastor Vargas (2014) show, from 2005 to 2013 the attendance rate went up from 57.3% to 70.8% in preschool, and from 70.6% to 80.3% in secondary school, while maintained its level around 91.3% in primary school. However, when comparing between urban and rural areas, the general picture for both preschool and secondary school changes dramatically; In 2012, for urban regions, the attendance rates to preschool and secondary school were 74% and 85.2% respectively, while in the rural regions were 63% for preschool and 69.9% for secondary school. By comparing between relative incomes the gaps persist; The high income group exhibits 17% higher attendance rates for preschool and 9% for secondary school. On the contrary, in both comparisons, between rural and urban regions, and high- and low-income groups, the attendance rates for primary school remain narrower: Less than 1% percent for the former and 5% for the latter dimension, highlighting the universal character of primary school in Peru.

[1] As Ortiz Chávez and Pastor Vargas (2014) note: On the first stage, the students take the first steps to self-identify themselves as individuals. During the second stage, the students start to train both their logical/mathematical reasoning and language skills. Between the third and fifth stages, the students consolidate their written and oral communication, and foster their mathematical analysis skills. From the sixth stage onward, abstract and reflexive reasoning is introduced. On the seventh stage, the emphasis is put on the autonomy of the communicative and moral expressions of the students.

Figure 1.2 Nationwide Number of Schools by Level (2004–2019).



Source: School Census, MinEdu. Cohorts are highlighted in gray and white. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

Finally, in terms of retention and completion rates, Ortiz Chávez and Pastor Vargas (2014) mention that one out of five children that starts primary school is unable to finish, while from those who complete primary education, 7.7% abandon the education system.

This general picture is consistent with the estimates of the National Household Survey (Instituto Nacional de Estadística e Informática [INEI], 2019), which shows that, as of 2018, 4.9% of the population aged 25 and over did not manage to study any level of education, 25.6% attended at least one year of primary education, 38.7% at least one year of secondary education and 30.8% at least one year of tertiary education (14.2% non-college and 16.6% college). Remarkably, among the population aged 25 and over, men register higher educational levels, mainly in secondary (8.2% difference) and tertiary education (3.1% difference), while women show a higher proportion in lower levels. For instance, in 2018, 28.3% of women had only primary education, compared to 22.8% for men. Furthermore, the population in urban areas has higher educational levels than their peers in rural areas. However, 51.8% of rural inhabitants finished primary education.

The impressive enrollment and attendance statistics have been achieved by a steady increase of schools across the country. As shown in fig. 1.2, between 2004 and 2015, the number of preschools almost doubled, corresponding to the period of analysis of this study. Still, as depicted by Ames (2016), despite the expansion of coverage, during the same period, paradoxically, there was a process of reduction and stagnation of spending on education. In fact, the budget allocated to education has decreased since 1970, so that it was not until 1999 that the investment levels of 1975 were reached (2.5% of GDP). Although in recent years there has been an attempt to increase investment in education, the spending per student still lags behind the Latin American average. About it, Gregosz et al. (2014) complement that, like its peers across the region, Peru experienced large improvements in primary and secondary school enrollment over the past two decades. However, Peru continues to face strong challenges in educational quality and equity, with most students in the country failing to meet basic learning standards measured by both national and international standardized test scores. In 2016, less than half of the students in Peru met learning standards for second grade in primary school in reading comprehension, while about one third met them in mathematics. Even more disheartening, only 14% of students who made it to second grade of secondary school achieved learning standards in reading, and only 11.5% in mathematics.

CHAPTER 2

Data

The data were retrieved from different Peruvian governmental institutions: The Ministry of Education [MinEdu], the Ministry of Economy and Finance [MEF], the Ministry of Transport and Communications [MTC], and the National Institute of Statistics and Informatics [INEI]. MinEdu provided the National School Enrollment Database [SIAGIE] between 2013-2015; the School Census from 2004–2019; the results of the Censual Student Evaluation [ECE] from 2007–2018 and its associated surveys in 2015, 2016, and 2018; and the National School Map from 2021. Information on socioeconomic characteristics and household structure were retrieved from the Household Targeting System [SISFOH] database from 2012–2013 administered by MEF. On the other hand, specific information about human settlements throughout the country was complemented with the National Directory of Human Settlements derived from the XII Population and Housing Census conducted by INEI and the Human Settlement Database from the MTC. An overview of the variables, their type and their sources is displayed in table 3.1.

Child and Family Identification

The main obstacle regarding the child and family identification is the systematic inconsistency between and within databases, even when national id numbers were implemented decades ago. That is especially important when considering that from almost 8 million enrollment registries recorded each year (compared to approximately 7.2 million school eligible Peruvian children between 5 and 17 years of age, according to INEI, 2018), only about 3.5 million different children could have been uniquely identified, which represents more than 50% data loss. Moreover, the first waves of ECE had no id number associated, increasing the matching burden and its associated uncertainty.[1]

Further, to assign a family to each child and verify the consistency of their personal information, I used the SISFOH; a rolling census designed to identify people or population groups in a situation of poverty, vulnerability, or exclusion, through a socioeconomic classification (not poor, poor or extremely poor), and assign and prioritize the access to public benefits and programs (Gob.pe, 2021). Therefore, SISFOH covers most of rural and a significant amount of urban households with data on socioeconomic characteristics, household composition, and labor market participation throughout the whole country and including almost 22 million individuals (approx. 74% of the total population).

Finally, in connection to the identification strategy, three assumptions used to identify and assign the children are noteworthy: (i) Since the information provided by SISFOH only covers 2012 and 2013, it is assumed that the observed family structure is held constant during the analysis period, which could be amply debatable on the light of evidence showing negative effects of family composition shocks on children's socio-emotional development and cognitive achievement (Lee & McLanahan, 2015). (ii) The available data does not help to determine with certainty explicit biological links between parents and siblings. Therefore, siblings are assumed to be children living "under the same roof" between 2012 and 2013, whereas the parents are assumed to be the reported household head and his/her partner. (iii) I proxied the household location area using the information of each test taken by any of the siblings and its associated school,

[1] That means that the matching procedures relied on a general rule on name matching. To my advantage, Hispanic tradition uses both last names —from father and mother side—, thus, rendering the name matching most robust to possible homonyms and duplicates. Further details are skipped given the length constraints of this document. However, the codes and the data assembling reports can be shared upon request.

reconstructing a "dynamic" history of family relocations, as I know with certainty when and where each test was administered. If reported schools were located in more than one area in one simultaneous year, I registered the family as living in the area belonging to the mode.

Preschool Tracking and Treatment Assignment

As previously highlighted, despite the ample reach of the administrative databases employed, one of the main challenges of this analysis is the consolidation and harmonization thereof. Currently, except from one subset of observations between 2015 and 2018, it is not possible to determine if any given identified child could access and effectively went to preschool when he/she was eligible. To that end and with the current means, it is only feasible to establish if a child had a preschool available close to home at a predefined age, what the literature commonly refers to as being assigned to a treatment. Nonetheless, as the concept is more imprecise, it is still difficult to assess it unambiguously. As a result, I performed three steps to verify the assignment to treatment for each identified child in a compelling way; (i) Even supposing that there is no preschool available in a town, —depending on the settlement patterns, the accessibility, and other local and geographical characteristics— one cannot overrule the fact that people (children) commute to get to work (school), which implies that it is necessary to determine a reasonable commuting area for preschoolers, where to verify availability. (ii) As explained in table 1.1, the preschool level in Peru comprises a regular and a community program, and thus, both options' availability has to be taken into account when estimating the effect. (iii) Because, supposedly, preschool education serves children between 3 and 5 years of age without any other further distinction, it is crucial to accurately ascertain how the eligibility rules condition the treatment assignment and account for it when determining the latter. I revisit each of the mentioned issues in the following subsections.

Commuting Areas

Peru is a land characterized by few high density cities and a vast number of small settlements with low population across the territory. In 2007, Guadalupe et al. (2017) show that out of 86,000 active settlements only 735 had at least 2,000 inhabitants.[2] As shown in fig. A.1, this trend has been consolidated in the last years, since, in 2017, out of 77,085 active settlements only 823 had at least 2,000 inhabitants. This difference is explained by two factors: (i) a rural-to-urban migration, resulting in the nominal decrease of active settlements (ii) an internal migration led by economic motives, deriving an increase of population density within the settlements (INEI, 2018, 2021). Such a settlement pattern has had a profound influence over the School Education System. Which, coupled with the persistently increasing enrollment rates observed during the last two decades, has turned the School Education System in the broadest social, organizational and governmental structure in Peru, not only due to the volume of the population it regularly serves —one out of every three Peruvians— but also because of its geographic scope (Guadalupe et al., 2017).[3]

[2] A settlement is defined as community residential unit or neighborhood, which can comprise political, social or civic organizations (Carswell, 2015). Active, on the other hand, refers to settlements with at least one permanent housing facility. For instance, Guadalupe et al. (2017) report 156 different settlements within Lima Metropolitan Area, meaning that 735 is an upper bound of the total number of cities or administrative areas.

[3] To compare the dimensions of the School Education System, Guadalupe et al. (2017) establish that there is at least one regular education program in 30,600 settlements: 19,900 with Preschool Level, 27,000 with Primary School Level, and 7,900 with Secondary School Level. Whereas the Healthcare System —the second biggest one— had in 2013: 511 hospitals, 15 specialized institutes, 2,096 healthcare centers, 7,124 health outposts, 3,963 medical offices and about 65,110 doctors. While, in 2014, there were only 1,459 police stations nationwide.

Unfortunately, although in the last two decades the Peruvian government has made several efforts to increase the overall data measuring intervals and their quality, the previously mentioned characteristics present a relevant challenge to the identification strategy, considering the poor quality of detailed local geographical information in Peru when comparing to international standards (Pineda Zumaran, 2016). Evidence of it is the existence of multiple, seemingly contradictory, and theoretical outdated standards and procedures to count and characterize human settlements across governmental institutions, which constitutes a burden to local planning practices and, in the end, hinders governance capabilities at the community level (Asmat Linares, 2020; Pineda Zumaran, 2016). For instance, there are two main geographic databases; one produced by INEI —and subsequently used and updated by MTC— and the other developed by MinEdu. The former is part of the National Population and Housing Census executed every ten years, being the last two in 2007 and 2017, respectively. The latter, instead, corresponds to the School Census and its associated National School Map, which is carried out every year and characterizes not only the continuing of the school services across the country but also counts the number of enrolled school children and some school characteristics, such as infrastructure and number of teachers.[4]

In that regard, there are three main problems: (i) As argued by Asmat Linares (2020), the Population and Housing Census does not take into account the provision of services (i.e., Healthcare and Schooling) in each of the settlements to define its hierarchy at the local level.[5] Instead, it relies solely upon permanent population levels, disregarding other information sources. (ii) There is no standardized harmonization between databases, meaning that not only settlements identifiers (name and id number) vary across sources, also the number of reported settlements in each district; a problem which is worsened by the frequency differences between both above-mentioned censuses given that population patterns inevitably change over time.[6] (iii) As shown in the introduction, rurality is an important dimension to explain school performance in Peru. However, following Ministerio de Educación (MinEdu, 2017), even within INEI sources there are different rural-urban classification standards. For example, there is one used in the Population and Housing Census (a settlement is considered urban if there are at least 100 contiguous housing units), and another used in the National Household Survey [ENAHO] (setting the previous threshold to 400 housing units or 2000 inhabitants). Moreover, the very concept of rurality is a gray zone, where a dichotomous variable cannot properly account for the whole continuum of accessibility, service provision and settlement patterns across the country (MinEdu, 2017).

Therefore, to correct for the previous mentioned issues, and be able to establish if a community within a commuting area has had or not a preschool service available, I used DBSCAN algorithm (Ester et al., 1996) on the Human Settlement Database of the most recent National School Map (2021), correcting by name hierarchies within districts. As Schubert et al. (2017) point out

4 This database have a clear advantage over the INEI one because; (i) It covers comparatively more territory with increased measuring intensity (yearly records). To help the reader have an idea, MinEdu database reports nationwide approx. 153,000 human settlements, while INEI only around 99,000. (ii) It is already harmonized with the school locations, since it is mainly used to track school services availability across the country.

5 More specifically, Asmat Linares (2020) discusses that it is important to recognize the settlements that, although having a low population volume (less than 150 inhabitants), do have education or health services, because they can generate a certain level of influence and attraction from —even larger— nearby populations that do not have such services.

6 Both Guadalupe et al. (2017) and Pineda Zumaran (2016) emphasize a higher order problem which condition the appropriateness of all information sources: The creation of new districts is not the result of planning or land-use planning processes, but rather to a logic of claim, vindication or access —legitimate or not— to resources, which, together with other factors, leads to a serious problem of territorial organization and management in Peru.

> DBSCAN uses a simple minimum density level estimation, based on a threshold for the number of neighbors, `minPts`, within the radius ϵ (with an arbitrary distance measure). Objects with more than `minPts` neighbors within this radius (including the query point) are considered to be a core point. The intuition of DBSCAN is to find those areas, which satisfy this minimum density, and which are separated by areas of lower density. For efficiency reasons, DBSCAN does not perform density estimation in-between points. Instead, all neighbors within the ϵ radius of a core point are considered to be part of the same cluster as the core point (called direct density reachable). If any of these neighbors is again a core point, their neighborhoods are transitively included (density reachable). Non-core points in this set are called border points, and all points within the same set are density connected. Points which are not density reachable from any core point are considered noise and do not belong to any cluster.

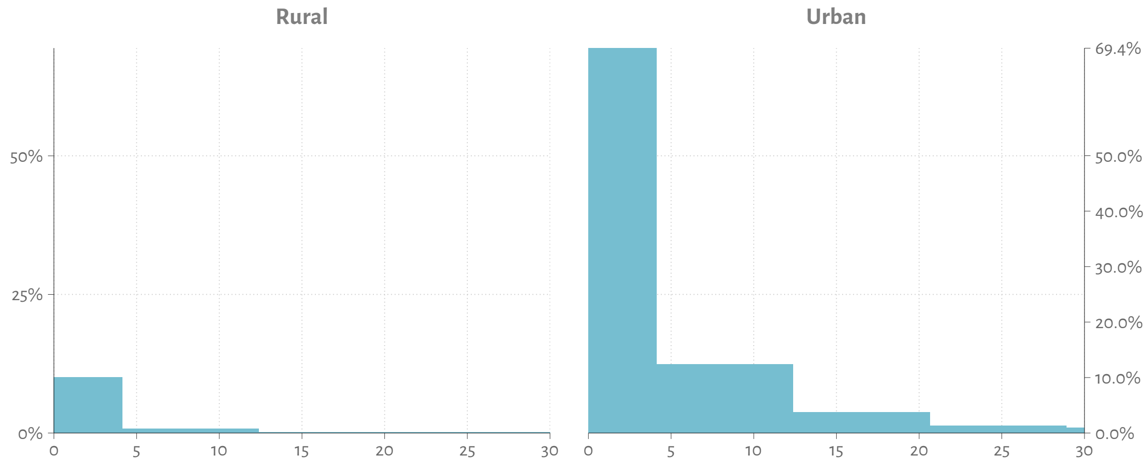
Therefore, DBSCAN offers multiple advantages for this particular application; (i) It can use arbitrary distance definitions, allowing the calculation of precise geographic positions and distances using the World Geodetic System Standard [1984]. (ii) It is based on transitive clustering, a feature that accounts for urban/rural continua and assigns all closely enough settlements to a unique location or community. That makes the algorithm robust to the high sparse settlement pattern of the Peruvian landscape. (iii) It only needs the geographic coordinates of the points to be executed, which is ideal in this case where no detailed nationwide geographic information exists. Still, all previous characteristics hold if and only if "reasonable" parametrizations are used. Consequently, it is necessary to define what constitutes a "reasonable" parametrization in the current setting.

First, since the objective is to find human settlements laying closely enough to be considered as the same unit, `minPts` was set to 2, the minimum possible detail. Nonetheless, ϵ cannot be set as trivially as `minPts` because it is well known, that even when DBSCAN performs well in small region accurate clustering, results are highly sensitive to small changes of ϵ (Yu et al., 2014). In this particular application, given the high number of points (around 150.000) and the documented slow performance of the algorithm (Schubert et al., 2017), it was not possible to optimize ϵ or use more sophisticated algorithms as HDBSCAN (McInnes et al., 2017). On that account, I used other sources to approximately determine the optimal commuting area for preschoolers in the Peru and merge human settlements based on several DBSCAN parametrizations.

Unfavorably to the current analysis, there is no detailed data on the accessibility characteristics at the local level in Peru, nor any study exploring the commuting patterns of school students at a national scale. To overcome that knowledge gap I resorted to four additional sources to help set a plausible commuting area for preschoolers (i) Salonen (2014) show that accessibility is greatly impacted by the rural or urban context, that "[t]he consideration of locally relevant travel modes is important, and the regionally specific transport network properties affect [how] distances and travel times need to be quantified." (ii) Therefore, although not being recommended due to its poor accuracy (Gao et al., 2019), using the average speeds reported in Gil and Read (2014)[7] and the traveling time and modes reported by children and parents in the Second Stage (2016) and Fourth Stage (2018) ECE surveys, I performed an approximated distance calculation discriminating between urban and rural contexts, yielding an average commuting distance of $3.936km$ [$Q_1 = 0.625km$, $Q_3 = 3.750$] and $3.909km$ [$Q_1 = 0.625km$, $Q_3 = 1.875$], respectively (see fig. 2.1 for additional details). (iii) As stated by Oficina de Infraestructura Educativa

[7] Which, even when other sources do not systematically list average speeds by mean of transport, the imputed values appear to be consistent with regional analyses in Latin America and local estimations in Lima, Peru (Rivas et al., 2019; Yachiyo Engineering, 2005; Yañez-Pagans et al., 2019)

Figure 2.1 Distribution of Reported Distance to Preschool.



Source: ECE Questionnaire to parents, Second Stage (2016) and Fourth Stage (2018). In the survey, parents disclosed information about the average commuting time and the mean of transport. Therefore, to get such a distribution, I imputed an average speed to each reported mean of transport and multiplied it with a normalized travel time. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

(2011), each newly built preschool has an influence area of $500m$ and $2000m$ for urban and rural zones, or between $15min$ and $30min$ of walking distance ($1.25km$ to $2.5km$ at an average speed of $5km/h$) (iv) Lastly, Rodríguez-Rodríguez et al. (2019) confirm the pattern of urban children commuting longer distances than their rural counterparts, as they found in Chile that, between the regions of Valparaíso (urban) and Easter Island (rural), 66% and 85% of the corresponding children travel an average distance of active commuting to school of less than $5km$. All in all, I chose a maximum distance of $4km$ to both urban and rural contexts as a "reasonable" commuting distance for the Peruvian case.[8]

Finally, I execute DBSCAN recursively using an ϵ from $250m$ to $2km$ with discrete jumps of $250m$ on each execution, totaling 8 clustering schemes with significantly different area distributions —ranging from $128,417$ to $9,497$ points labeled as noise, see fig. 2.3 for more details—, and selecting only those clusters whose polygons (adding a buffer zone of $1km$ from all borders of the generated polygon) have an area less or equal to $4\pi km^2$ (or a circle with $2km$ radius) in the widest possible parametrization.[9] Then, all points are merged if they belong to the same cluster based on the area-optimized cluster schemes, and their corresponding generated areas are calculated again to verify consistency. In addition, to account for position measuring errors, in the merging process I grouped all human settlements having the exact name and belonging to the same district, as long as such a constraint generated clusters with a definitive area less than $100km^2$. The resulting area by cluster distribution is shown in fig. 2.3.

Preschool Programs

Guerrero and Demarini (2016) reveal that, in spite of the universal status of the primary school in Peru from late 1990s onward, and from its public adoption in 1972 until

8 If judging by the variation of treatment vs control observations, the assignment seems to be successful. I had access to a survey were children were explicitly asked if they had attended preschool. While comparing the reported treatment and the estimated treatment assignment, the proportion remains fairly similar (approx. 0.09% for control units), which gives confidence about the approximation through commuting distances. Nonetheless, as no effect was credible estimated, I do not show the difference between samples.

9 For example, if at $\epsilon = 500m$ a cluster polygon has an area of $2\pi km^2$, at $\epsilon = 1km$ it has an area of $3\pi km^2$, and at $\epsilon = 1.25km$ an area of $5\pi km^2$; I choose the cluster generated by $\epsilon = 1km$ for that particular point.

mid 2010's, preschool education access was largely insufficient.[10] In that respect, and perhaps recognizing the supply limitations, even when along its adoption it was also acknowledged the crucial role of preschool to foster the development of the children, the Peruvian government did not enact preschool education as compulsory until the current "General Education Bill" (2004) sanctioned otherwise. In turn, the bill of 1972 emphasized that the communal and familiar engagement is key to promote children's integral well-being, and implemented shortly after regular and community preschool programs, as well as parenting workshops across the country (Dirección Nacional de Educación Inicial y Primaria [DINEIP], 2003).[11]

Nonetheless, the number of newly-established regular and community preschools did not increase steadily after the 70's. In 1996, when the number of community preschools were near to 23,000 and comprised around 40% of the total public preschool supply, a series of government reforms —aiming to reduce the public spending and increase its efficiency amid the trade liberalization— reorganized the education sector and cut the investment, specially the one funding the community programs. This led to a stepped reduction of quality, infrastructure and enrollment rates, as well as an interruption of the establishment of new programs (Llanos Zuloaga, 2016). It is only after 2002, when the National Plan of Action for Children and Adolescents [PNAIA] was launched to frame the central government's public policies in favor of children and adolescents and together with the "General Education Bill" (2004) set the legal and bureaucratic basis to scale-up the preschool education at a national level, making possible the doubling of the existing preschools in less than 15 years (Guerrero & Demarini, 2016).

Currently, the available preschool education programs for children between 3 and 5 years in Peru comprise two kinds of institutions; (i) Regular *Jardín*, which corresponds to the regular program, follows the official curriculum and is carried out in an established and dedicated building, serving between 20 and 25 children in both urban and rural settlements. (ii) Community *Pronoei*, which corresponds to the non-regular program, does not follow the curriculum but has a dedicated room —whether rented or owned—, where to carry on the planned activities. In *Pronoei*, a teacher implements pedagogical workshops with the participating children and promotes good parenting practices and a healthy and safe environment for children within the community.

The School Census administered by MinEdu provides the data on school characteristics, including location and opening (closing) of each of the programs.[12] Nonetheless, before 2011, the community preschool programs were not included in the regular census. Instead, they were registered in a different section using different identification codes, which are not codifiable into the same standard. That posed a problem even within MinEdu, as, to be able to determine the date of creation of each of the community programs, MinEdu had to carry out extra surveys. Unfortunately, even though the

[10] In fact, Guerrero and Demarini (2016) argue that preschool education in Peru has much more history, dating back to 1892. However, it is only after the General Education Bill N. 19326 in 1972, when it is officially recognized as a school level aimed to serve children between zero and five years of age, and nationwide efforts to broaden its coverage formally began.

[11] The community preschools stemmed out as a institutionalization of the *Wawa wasi*/*Wawa Uta* programs (Children's house in Quechua and Aymara respectively). Originally developed by Caritas Internationalis in early 1960s to supply children with nourishment, healthcare, and recreational and educational activities, after the Education Bill in 1972, MinEdu adopted those programs to expand preschool education to rural areas. At the beginning they were known as "Proyecto Experimental de Educación Inicial No Escolarizada" [*Propedeine*] but were later renamed as "Programa No Escolarizado de Educación Inicial" [*Pronoei*] (Guerrero & Demarini, 2016).

[12] MinEdu (2017) recognizes the lack of robustness of the database, acknowledging that education services are not grouped according to their membership in an actual educational institution. Therefore, it is not feasible to keep track of the school's physical characteristics or available equipment, rendering a further analysis of the differences between the programs impossible.

dates can be retrieved, such a situation introduces uncertainty while determining the availability of the programs because of possible systematic underreporting of the actual status, biased towards non-availability, and possibly decreasing the estimated effect, if any exists.

Elegibility

Even though when the "General Education Bill" (2004) clearly states that "The age of entry to the different school education levels will be fixed in accordance with the flexibility principle, prior to an appropriate assessment", MinEdu established that enrollment is only possible if the child complies with the predetermined age of entry before each March 31st —which roughly marks the start of the school year— to ensure that the children start primary school at 6 years of age, a rule which is also used to determine eligibility for preschool (Guadalupe et al., 2017). Basically, following Kagan (1990), this cut-off rests on the presumption that development precedes learning. Hence, it is necessary (and reasonable) to assure a minimum level of cognitive and non-cognitive development from which to build further skills through formal learning.

However, as shown by Morales (2020), such a seemingly innocuous cut-off has had demonstrated effects on learning outcomes, which operate through two main mechanisms: (i) The cut-off generates exogenous variation deriving in comparison problems; across cohorts, as relatively same age children belong to different grades; and within cohorts, as there could be children one year older (younger) belonging to the same cohort. (ii) In general, the composite effect between school entrance age, chronological age, and years of schooling is a mathematical identity, and cannot be disentangled. In terms of academic achievement, Morales (2020) concludes that "the empirical literature on school-entry age effects on children's academic achievement (...) [show a] positive association between standardized test scores and being older at school entry" and that "a larger effect is predicted for [preschool] and elementary school grade levels, and it seems to persist through at least eighth grade".

Nonetheless, Guadalupe et al. (2017) report that, in Peru, there is systematic misreporting of enrollment ages, in addition to increasing early-enrollment despite the regulations.[13] A fact that is evident in the SIAGIE database, where 96% of the observations have inconsistent birth-dates with respect to the taken test. As discussed before, that is notably problematic given the significant effects produced by the cut-off, which for tests specifically designed to account for learning outcomes on curricula —such as ECE— could exert serious bias, with the aggravating problem that I cannot calculate age-at-test deviations reliably. Consequently, to correct for systematic birth-date errors with the available data, I assumed that month and day of birth are reported correctly, and created a birth-date eligibility cut-off indicator variable (being 1 if the month and day of birth is equal or before March, 31st; 0 otherwise) as the best approximation to account for eligibility age concerns.[14]

Poverty Measurement

As argued before, socioeconomic status seems to have a leading role in the determination of cognitive development, and thus, of learning outcomes too. It appears not to be only related through a direct channel, but also as an indirect determinant of other

[13] The fraction of 5-year-old children enrolled in first grade of primary education compared to the total enrolled children was around 47% in 2006.

[14] However, as pointed out by Morales (2020), the Peruvian institutional context provides more complications because the cut-off is not always enforced, in addition to several cut-off changes across years. I acknowledge the limitations and leave their explorations for later works.

ecological factors like "family income composition, and social support; parental level of education, occupation, mental health, parenting style, and parent-child interactions; housing conditions, quality of home and school environments, early childhood program attendance; neighborhood resources; and health and nutrition" (Segretin et al., 2016). Yet, the evidence for Latin America is still scarce. One explanation thereof is the plethora of poverty and cognitive development definitions, as well as instruments to measure them, which inevitably hinders the establishment of systematic review and comparison frameworks. Moreover, following Bonal (2007), from a public policy perspective, poverty has been fought through extensive education investments but, such a perspective, often adopted in governmental institutions, fails to recognize that, beyond fostering social mobility, at the macro level, the quality of education is also determined by overall socioeconomic conditions.

In that regard, Duarte et al. (2010) shed light on the relation between socioeconomic status —measured using parent education level variables, characteristics of housing, access to public services, and family access to cultural assets (e.g., books at home)— and SERCE, a standardized test administered by the Latin-American Laboratory for Assessment of the Quality of Education [LLECE] and UNESCO measuring student achievement in relation to curriculum objectives across Latin America (Center for Global Education Monitoring, 2014). Duarte et al. (2010) find not only that there is a positive relation between the socioeconomic index and the test scores, but also that the relation appear to be stronger across schools than within schools, confirming the segregation patterns already documented by the literature (Torche, 2014).

Therefore, to account for the heterogeneity generated by different socio-economic standings, it is crucial to be able to measure it reliably. For that, I recreate the Socioeconomic Index [ISE] developed by MinEdu, using variables on parental education; and construction materials, basic utilities, assets, and other services at home (Ministerio de Educación, 2017). ISE index corresponds to a socio-economic status [SSE] (wealth) index offering various advantages over other kinds of measures because; (i) Since it is based on surveys, it uses household characteristics that can be measured directly by the surveyors' team, in contrast to self-reported income or expenditure information which is subject to under-reporting and other kinds of measurement errors (Córdova, 2009). (ii) It captures non-monetary income streams, especially important for poor families who may have income in kind. (iii) Although it is unsuitable in certain applications where income shocks are important, wealth indexes can measure long-standing family economic conditions, being robust to transitory macroeconomic or regional shocks, and being able to represent wealth distribution dynamics across the population (Kabudula et al., 2017).[15] (iv) It shows good invariance properties across students covariates and appropriate classifying power of socio-economic profiles in developing contexts (Bofah & Hannula, 2017). (v) Wealth indexes are proved to be a valid and reliable well-being measure across international contexts, being correlated with long established measures like the Human Development Index or poverty headcount ratios (Smits & Steendijk, 2015). In summary, as Bofah and Hannula (2017) assert:

> It can be concluded (...) that students' reported household resources provide comprehensive data on family background. (...) [T]he approach considered here will serve as a practical guide for educational researchers seeking to construct a reliable SES measure in low-income societies and in studying educational inequalities related to family background when using large scale

[15] However, one might assume that short-run family income dynamics are not important to determine the educational outcomes, Lehti et al. (2019) show that, even when unemployment does not have cumulative effects regarding a child's age, at the transitional education periods, it actually may have a determinant role, especially when transitioning to secondary education.

international studies.

The information on household resources was collected on a sample basis employing a questionnaire to parents and students in the ECE, and the data stored in SISFOH. Although I could try to reconstruct the index as a time-series, given that the ECE questionnaires were only administered to a reduced sample, and there is only data for all families between 2012–2013, following Vyas and Kumaranayake (2006), I calculate the index for the whole pooled sample and then average it across families to get a cross-sectional index. In short, the ISE index is an aggregated standardized deviation weighted by the first-dimension PCA loadings of the maximum education level attained by one of the parents and three sub-indices; materials, assets, and other services at home. Each sub-index also corresponds as aggregated standardized deviation weighted by the first-dimension PCA loadings of different items ranked by well-being levels. In the end, four groups —high, medium, low, and extremely low— were defined at the percentiles 35, 60, and 85. Nonetheless, I used the continuous scale to account for smaller differences, although it precludes its direct interpretation.[16]

Learning Outcomes and Standardized Tests

Along the efforts to universalize primary education, while considering that quality is an integral component of an equitable education system, MinEdu decided to implement nationwide standardized tests to measure the learning outcomes of primary school students, enabling the accountability of the education system at school, district, and province levels. The test was administered for the first time to second-stage students in 2007. Later, in 2015, the test included sixth-stage students. Finally, in 2016, fourth-stage students started to be evaluated.[17] Unfortunately, after months of political unrest, a generalized school teachers' strike, and the strongest "El Niño" event recorded in the last 20 years (Rodríguez-Morata et al., 2019), MinEdu canceled the planned tests for 2017 since "the variability of class recovery no longer provided the basic conditions for the implementation of a standardized assessment of learning outcomes" ("Ministerial Decree No. 529-2017-MinEdu", 2017).[18] Therefore, considering that the test scores constitute the only outcome variable of the present analysis, and they are objected to significant variability and uncertainty —as exemplified before— I will revisit the basic concepts of test comparability and its threats. Later, I will explain how ECE's methodology works and discuss why it is an appropriate measure to account for the effect of preschool education.

In general, assessing the validity of standardized tests is a rich research area of education and psychometric research. However, as González and Gempp (2021) argue, the relationship between validity and comparability is not entirely clarified, even when comparability is a key component of any education assessment method. According to American Educational Research Association et al. (AERA et al., 2014), validity is "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. [Therefore, it is] the most fundamental consideration in developing tests and evaluating tests". Whereas comparability is generally defined as the degree to which individuals with the same score possessed the same level of proficiency with respect to the domain

16 One issue pointed out by Contreras et al. (2019) remains a concern. Since the probability of observing a switcher family increases with respect to the household size —a fact that is taken into account in the calculation of the propensity score—, the relative socioeconomic condition could also be worsened by the overcrowding effects identified by the authors. That, in turn, could lead to a systematic underestimation of the treatment effect, assuming the socioeconomic status is not explicitly controlling for such phenomenon.

17 Although the fourth-stage test was first administered in 2012, it targeted bilingual and multicultural schools only. Thus, acknowledging comparability threats, its scores will not be used in the present work.

18 This explains why in fig. 2.2 and all other figures in appendix A, there are no score data for 2017.

a test was intended to measure, even if those scores come from measurements taken at different times, in different places, through different modes, with different items, or administered to different sub-populations (González & Gempp, 2021). Consequently, if the test is intended to reflect a process over time, to enable accountability, or more generally to be used as guidance for different interventions or policy-making decisions, comparability is then a fundamental element of the validity of a test. In that regard, as Berman et al. (2020) analyze, both validity and comparability face several threats: (i) Measurement errors (ii) Inaccuracy between what the instrument measures and the actual abilities intended to be measured (iii) Systematic differences in scores among students with equal proficiency with respect to what the tests are designed to measure.

Fortunately, around 2006, when ECE was being conceived, the previously mentioned concerns were taken into account, and through extensive use of the then-available literature, MinEdu (2016) developed a test that seeks to: (i) provide information on the learning achievements of Peruvian students in mathematics and communication for all stages, and history, geography, and economics for the sixth stage; (ii) compare results over time and report the evolution of learning outcomes. To achieve that, the ECE uses a staged procedure that considers and ensures validity and comparability —as discussed before— in every phase. First, a group of experts define what constitutes the evaluated skills and formulate the items, an assessment based on the national curriculum guidelines established by MinEdu. Then, using the Rasch model for dichotomous items constructs a ranking scale of items according to their difficulty, assuming that correctly answering an item is a probabilistic function—.[19] Such a scale presupposes that the greater the ability, the more likely the person will be able to answer correctly and then, estimates an interval index of skill. Although the exact psychometric traits of the test are out of the scope of this analysis, the following considerations are noteworthy: (i) The test scores are reported on an interval scale, which means that standard mathematical operations are meaningful. (ii) Albeit the scale is arbitrary, repeatedly using some items across years allows the inter-temporal calibration of the scale with respect to a certain year and compares results across cohorts.[20] (iii) The modeling assumptions and the functional forms used ensure that the test is uni-modal —allowing the use of boxplots to represent distributions, as in fig. 2.2—, and assures the finite variance of the scores (crucial trait for the statistical assumptions described in chapter 3). Finally, (i), (ii), and (iii) imply that it is possible to analyze the dynamic effects of preschool by directly comparing the (standardized) magnitudes and standard errors of the estimated effects.

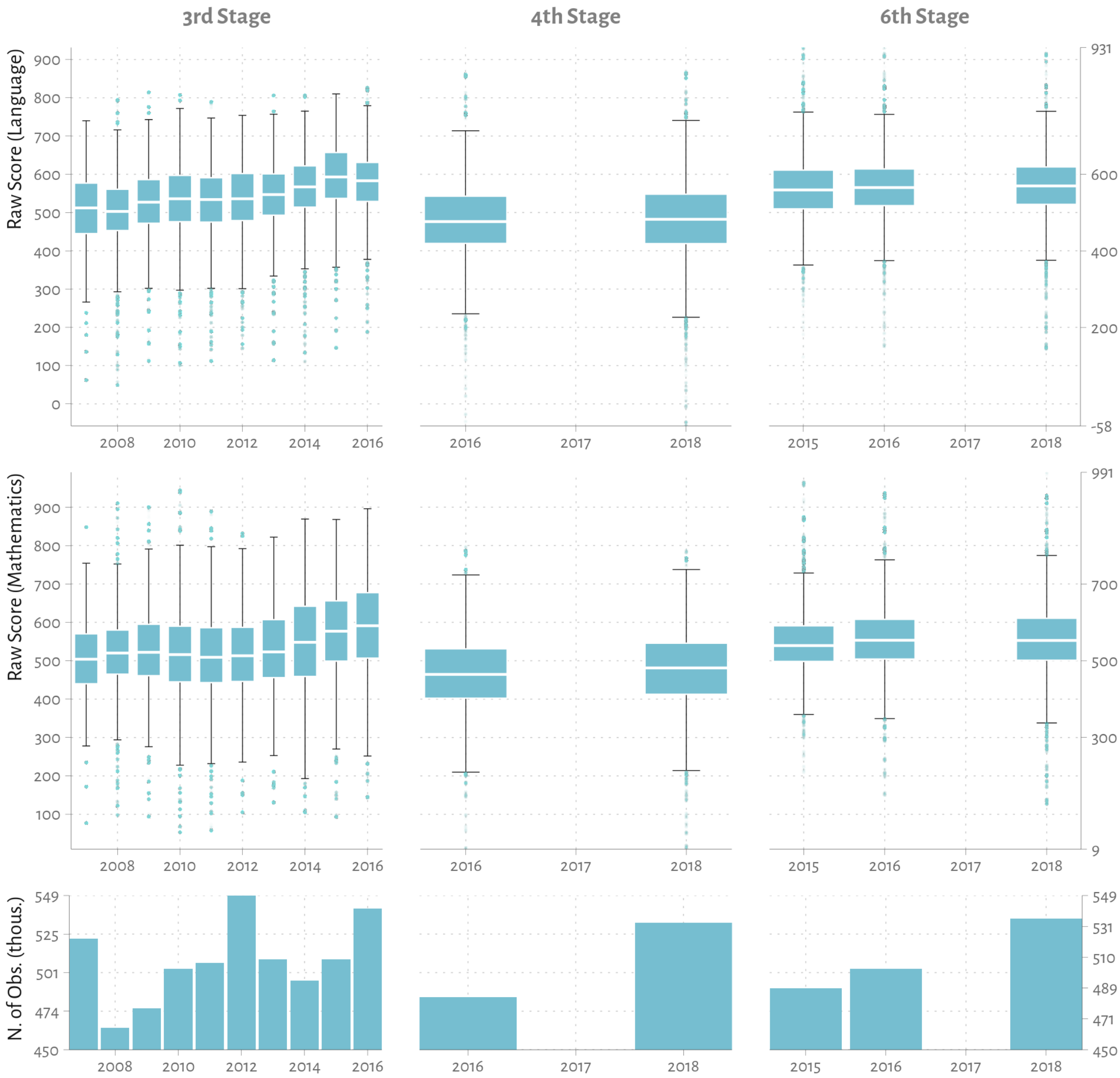
In consequence, and concerning its goals, ECE covers all schools registered in SIAGIE with more than 5 students. However, given that the accessibility conditions of several regions preclude the continued enforcement of quality controls,[21] the test has two versions; a more rigorous control sample, which evaluates a representative subset of the school distribution of the country (with respect to rurality, gender composition, teaching conditions, and school status; public or private), and the censual test, which covers everything else left. Therefore, considering that restricting the analysis to the control test would exacerbate external validity concerns, given that the distributions could systematically differ on unobservable characteristics, I will use the entire sample reported by ECE and leave such consideration for future works.

[19] As highlighted by MinEdu (2016), the Rasch methodology models each item as the interaction between an item and a person, establishing that the probability of a person's correctly answering an item is the difference between the person's measure of skill or latent ability and the relative difficulty of an item.

[20] Following MinEdu (2016), all tests are standardized with respect to the second stage test (control sample) administered in 2007, which has a mean equal to 500 and a standard deviation of 100.

[21] Because the tests abandon the safety circuit some days before their actual administration, and, for example, it cannot be assured that no student had prior access to the sheets.

Figure 2.2 Test Scores [ECE] (2007–2018).



Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

All in all, of the total identified students, on average, 95% are effectively evaluated each year, corresponding to approx. 500,000 students in each participating stage. Moreover, as shown in fig. 2.2, there has been significant improvement over the years. However: (i) the rates are not constant, showing stronger changes only at the end of the period. (ii) Despite progress, the proportion of students who achieve the minimum achievement level is modest, with more than 50% fail in language, whereas approximately 75% do not achieve it in mathematics. (iii) It is possible, as pointed out by Guadalupe (2015), that using ECE as an accountability device for schools, programs, and regional governments, could have led to an "optimizing" behavior with respect to the test, with anecdotal evidence showing that schools have explicitly offered courses and workshops to prepare students before taking the ECE. That, in turn, would necessarily imply that the instrument is losing validity with time and that observed improvements do not reflect the actual progress of the education quality but just an adjusting process from the agents' side.

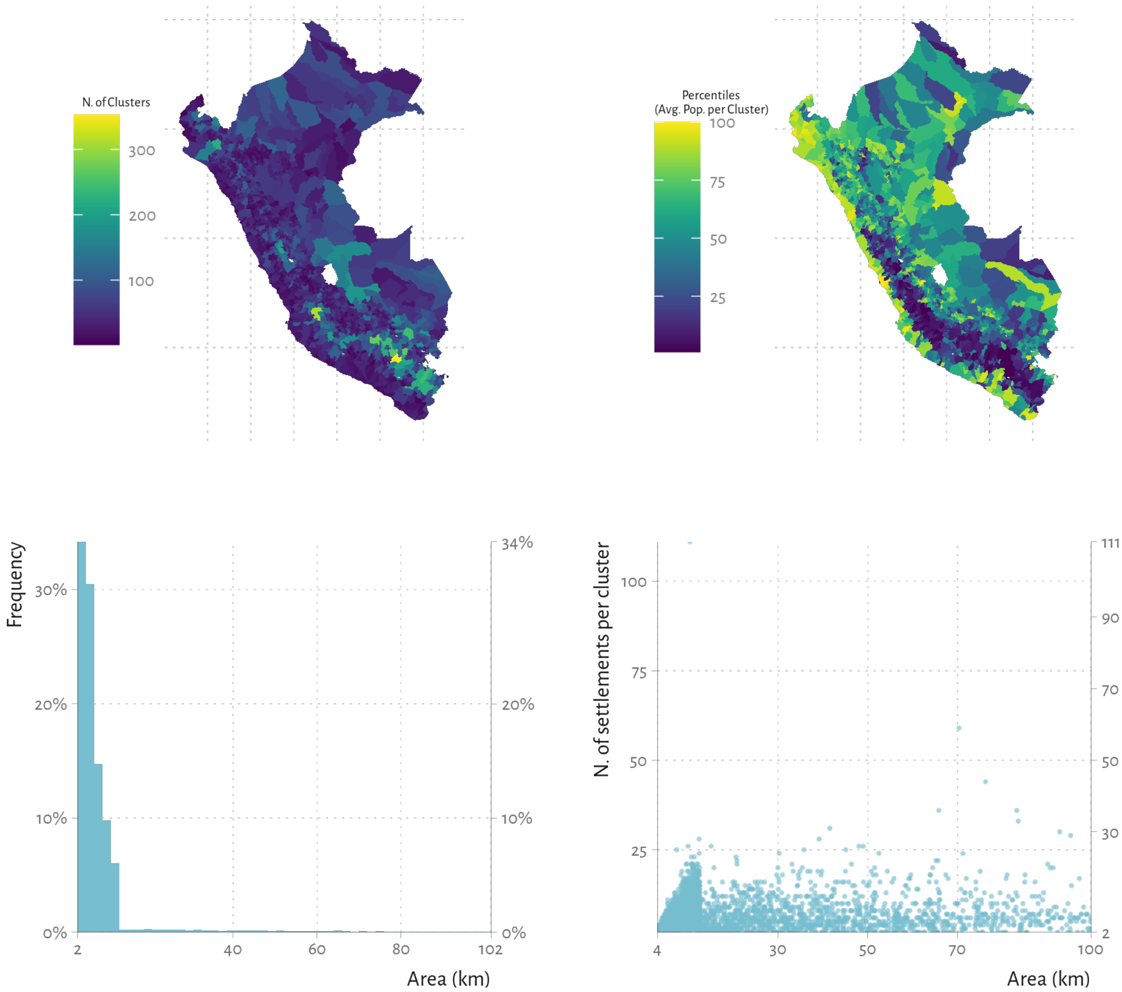
Therefore, I use cohort dummies in the estimation specification to control for time effects, as following: [2002, 2005], (2005, 2010], (2010, 2013], (2013, 2016]. Cohorts are defined as the year when children were eligible to attend preschool and were chosen considering (i) the available sample sizes and (ii) the school system's historical context. As described before: cohort 1 roughly corresponds to the pre-reform period, cohort 2 captures the first roll-out period, when both the ECE and preschool programs were being adjusted, whereas cohort 3 stands for the fastest-growing period. Finally, cohort 4 represents the maturation period, both of the ECE and the national preschool education supply (see fig. 1.2)

On the other hand, as shown in the appendix A, the observed gaps in enrollment, attendance, and completion rates —described in chapter 1— are also observed in ECE; between rural and urban regions (benefiting urban), and between private and public institutions (benefiting private). Such gaps indicate no easing over the cohorts and remain a structural feature of the School Education in Peru. Interestingly, even when the medians between boys and girls differ (benefiting boys), in terms of distributions, the difference is negligible across tests and cohorts —especially in mathematics—. The same stylized fact is confirmed by Guadalupe et al. (2015) using the control sample only, stressing the importance of controlling for those covariates in this particular design.

Other Relevant Covariates

Finally, I add some control variables which, according to the literature, have significant effects on learning outcomes, such as birth order indicator (1 if first-child, 0 otherwise) (Barclay et al., 2021; De Haan et al., 2014), family size (D. L. Miller et al., 2019), an indicator for access to social programs and other for formal healthcare (Burkett et al., 2020; León Mendoza, 2019), an indicator of the main language at home (1 if different to Spanish, 0 otherwise) (Navarro et al., 2018), and a uniparental family indicator (Saracostti et al., 2019).

Figure 2.3 Settlment Clustering: DBSCAN Outcomes at District-Level.



[*i*] Upper-left: Depicts the number of identified clusters within each district. However, in reality, the clusters were made without taking into account administrative borders. Thus, the plot corresponds to a simplification since there were more than $150,000$ single human settlements across the land. [*ii*] Upper-right: Shows the belonging of each district to the percentiles of the population density per cluster distribution, i.e., the total population living in a district divided by the number of clusters in the same district in 2017 [$min = 0, Q_1 = 60.05, Q_2 = 131.82, Q_3 = 302.77$ and $max = 75359$]. Since the urban areas had bigger districts, with more population, showing the actual values would hinder the analysis given the huge differences. [*iii*] Lower-left: Draws the area distribution of clusters with more than one settlement ($31,393$ of $67,867$ total clusters, or the 46%). [*iv*] Lower-right: shows the correlation between area and number of settlements per cluster. First, it is evident the clustering at $4\pi km$, since it was the main restriction. All other points beyond the limit correspond to settlements clustered in the same group because of having the exact same name. *The right vertical axis is meant to help reading the graph, as it mirrors the left vertical axis.*.

CHAPTER 3
Identification Strategy

To estimate the effect of attending preschool on learning outcomes, I exploit the exogenous variation produced by the public expansion of preschool institutions carried out by The Access to Preschool Program launched in 2010, which in words of Majerowicz Nieto (2019) "was [designed to] close large rural-urban gaps in access to preschool education, as well as improve overall enrollment rates which were low, (...) [establishing] —from 2011 to 2015— over 30,000 [new] preschools (...) all over Peru.". The expansion was entirely financed and designed by MinEdu, but the decision was taken by each Regional School Board (Unidad de Gestión Escolar Local[1] [UGEL]). Thereon, Majerowicz Nieto (2019) sheds light on how the process was executed:

> [First,] UGELs selected towns with potential unmet demand (defined as the number of preschool-age children without access to preschool), and generated "demand studies" where teams would visit prioritized towns and do a census of the student population by going door to door and identifying all preschool or soon-to-be preschool aged children in the town. [Then,] each UGEL prioritized towns according to unmet demand using the results from these studies, and the top towns were selected to receive a school. The type of school that got built was determined by the number of preschool-age children in a given town, according to the following rule: (i) 15 or more students; [a] *Jardín*–Regular preschool [was built] with a designated classroom and at least one fully trained government teacher (ii) 8-14 students; [a] *Pronoei*–Community preschool [was established,] where [a] local mother runs the school under occasional supervision of a teacher coordinator. (iii) 7 or fewer, [a] household visit [program was carried out,] where a trained teacher visits parents once in a while to teach them how to play with the child (p. 50).

Therefore, given that there was no explicit randomization on the treatment assignment, to credibly estimate the impact of the preschool expansion on learning outcomes, I have to discuss both external and internal validity concerns regarding non-random selection, as this would constitute an ex-post single difference design (White & Sabarwal, 2014). In effect, even when there are time-varying data available (i.e., across years and school stages), they only allow estimating a single difference design because; (i) The treatment is assigned before measuring the outcome variables since the standardized tests are administered in primary and secondary school, which excludes any pre-treatment analysis (ii) While there are various cohorts, by design, each measurement is unique in relative (developmental stage) time —except if children were retained—, which, together with (i) effectively implies a cross-sectional design. Moreover, if one considers using a difference in differences design setting families as the treatment units, two further problems arise; (i) As the matching between databases was not perfect, the end panel is heavily unbalanced on calendar time (ii) Individual-level variables such as sex have to be dropped.

[1] Although the Spanish name refers to each School Board as "local", in reality, each UGEL oversees a sizeable group of schools across one or multiple districts. Nationwide there is substantial variation between UGEls, although, on average, each UGEL serves between 8 and 9 districts (MinEdu, 2017). For instance, currently two UGELs manage 33 districts, while another four UGELs manage only one. Therefore, it is more appropriate to denominate them as "regional", given their characteristics, and to keep a uniform conceptualization throughout the text.

Internal Validity

In terms of internal validity, just comparing differences in means between the treated and non-treated groups would not yield causal estimates, as families who sent their children to preschool and those who did not have systematic differences that simultaneously impact the learning outcomes of their children. For example, as shown in the introduction, rurality, poverty, and the public status of the school are relevant dimensions to explain test scores' performance, in addition to other unobservable traits such as parenting practices and styles. Moreover, the previously described decision process implied that the establishment of new preschools was not random. Instead, each UGEL was autonomous in choosing which towns were to be included in the demand studies, which in turn were later used to determine where new preschools were going to be built. Following Majerowicz Nieto (2019), "In fact, in conversations with the office of Access to Preschool, it was difficult to determine clear criteria for how these towns were chosen other than some general 'suspicion' (often based on local knowledge) that the there was unmet demand there". Therefore, comparing between areas with differential exposure to the program would also yield biased estimates, if the assignment process is correlated at the same time with other area aspects that condition learning outcomes (e.g., local governance, accessibility, service provision, etc.)

In that regard, literature commonly uses family fixed effects to solve the previously mentioned selection issues, exploiting the variation in the exposure of siblings to the program caused by its timed roll-out. Such a specification allows the within-family differences to control for unobservable variables common to all siblings while across family differences to control for changes in observable characteristics on both individual and family levels (see Barclay et al., 2021; Deming, 2009; Pages et al., 2020; Prime et al., 2017 and Schlotter, 2011 for related examples). Unfortunately, whilst the family fixed effect design can handle unobserved family characteristics, given data unavailability, it is not able to take into account unobserved traits at the individual level such as; innate cognitive and non-cognitive ability, parental favoritism as well as nutritional and health conditions, all of them being relevant unobserved variables for this particular application.

External Validity

Still, even when —in theory— internal validity concerns could be taken care of by the inclusion of family fixed effects and a rich set of control variables at both family and individual level, as argued by D. L. Miller et al. (2019), external validity concerns are not appropriately handled with the sole use of fixed effects. As a fact, standard fixed effects induce a special type of non-random selection in estimation, called "selection into identification", if treatment "switcher" groups are (i) a subset of the sample and (ii) systematically different than the overall population. For this particular case, selection into identification is very relevant as the switchers are those families with varied treatment exposure among siblings, possibly inducing bias inasmuch as the effect of the treatment can only be identified on those groups. More specifically, D. L. Miller et al. (2019) show that; "in the presence of heterogeneous treatment effects, [selection into identification] causes [fixed effects] to deviate from the [Average Treatment Effect (ATE)]" because family fixed effects induce the use of a reduced sample —using only those that have exposure variation—, and because such a loss of sample variation is systematically related to relevant observables —capturing asymmetrically underlying subgroups—. To solve that precise concern, D. L. Miller et al. (2019) propose to; (i) estimate a "reweighting-on-observables method that can be used to recover the ATE for the overall population or target populations (such as program participants)", (ii) report "the sample size when limited to switcher families and quantify the contribution of residual switchers",[2] and (iii) "show the balance of covariates across switcher and non-switcher families".

Estimating Equation

Following Deming (2009), the identification assumption is that (∗) *Selection into preschool among members of the same family is uncorrelated with the unobservable determinants of outcomes*. The estimating equation is therefore:

$$Y_{ij} = \alpha_j + \beta D_{ij} + \delta D_{ij} \times C_i + \gamma X_j + \epsilon_{ij} \tag{3.1}$$

where Y is the (language or mathematics) score of the (second, fourth or sixth stage) standardized test, D is a dummy that takes value 1 if a preschool was available in the area of student i from family j when she/he was 5 years old, 0 otherwise; α_j is a family fixed effect; C_i is a vector of individual-level controls (i.e., sex; cohort; first-born, birth-date eligibility cut-off, public-school and single teacher school indicators), and X_j is a vector of controls that vary across families (i.e., socioeconomic status, household size; and rurality, participation in social programs, main language at home, formal healthcare access, and uni-parental family indicators). As suggested by D. L. Miller et al. (2019), the interaction term $D_i \times C_i$ is added to get an estimate for the ATE for the residual switcher groups: $\hat{\gamma} \times \overline{C_i} + \hat{\beta}$. However, to use this specification and retrieve $\hat{\gamma}$, the effect of individual covariates C_i is assumed to be constant and homogeneous across families. Then, the weighting scheme remains unchanged, and the differences in means are meaningful. Standard errors are clustered at the family level.

The equation is estimated using the equivalence between a within-group matching estimator and a weighted linear group fixed effects regression model established by Imai and Kim (2019), which also allows addressing external validity concerns through selection into identification corrected weights as developed by D. L. Miller et al. (2019). Therefore, the estimated causal parameter corresponds to the Average Treatment Effect (ATE) $[\beta = E(Y_{ij}(1) - Y_{ij}(0)|T_j = 1)]$, where T_j is a predefined target group from the population, taking into account both internal and external validity issues as previously described. Nevertheless, it is important to highlight that in the main sample, the treatment corresponds to the *availability* of preschool regardless of whether the students attended preschool or not, thus, representing an intent-to-treat estimate (ITT). For a subset of observations from 2015 to 2016, I have survey data explicitly asking parents or children —if they were taking the corresponding sixth stage exam— whether they attended preschool or not. In that case, such a group can be regarded as compliers, and therefore I can recover the ATE. I report both results in chapter 4.

Estimation Framework

Imai and Kim (2019) develop a general matching method for causal inference which corrects the bias of traditional fixed effect estimators through a non-parametric transformation, proving that "a within-group matching estimator is equivalent to a weighted linear group fixed effects regression model", where its non-parametric specification allows different forms of treatment heterogeneity.[3] Consequently, to consistently estimate the

[2] Residual switchers refer to the case where covariates C_i —those that vary across i units within a group— induce variation in the residualized (differenced) group treatment, even if there is no within-group variation in the assignment-to-treatment D_i

[3] Indeed, Chernozhukov et al. (2013) establish that the linear group (unit) fixed effects estimator fails in both cross-sectional and panel settings to consistently estimate the ATE unless either the within-group ATE or the within-group proportion of treated observations is constant across units (homogeneity). Such a proposition also covers stratified randomized experiments. Nevertheless, since that discussion is not the main objective of this text, I strongly recommend to follow Imai and Kim (2019) and D. L. Miller et al. (2019) for further details.

ATE using Imai and Kim (2019) estimator together with the reweighting-on-observables method suggested by D. L. Miller et al. (2019), the following assumptions —in addition to assumption ($*$)— have to be fulfilled: (*i*) There are no spillover effects both across and within groups. (*ii*) There exists no unobserved group- or individual-varying confounder $\epsilon_{ij} \perp\!\!\!\perp D_{ij}|C_i, \alpha_j$ (*iii*) Conditional on observables, the treatment effect is independent of a unit's switching or target status. (*iv*) $\mathbb{E}(Y_{ij}^2) < \infty$. (*v*) The calculated propensity scores are correctly defined.[4] (*vi*) There is a positive probability of being a switcher for each combination of group-level variables in the target group. Then, as shown by D. L. Miller et al. (2019), the family fixed effects consistent ATE estimator for the predefined target population t corresponds to:

$$\hat{\beta}^t = \frac{1}{\sum_i \mathbf{1}(S_{j(i)} = 1)} \sum_{i|S_{j(i)}} \hat{w}^t_{j(i)} \times \hat{\beta}_{j,FE} \quad \text{and} \quad \hat{w}^t_{j(i)} = \frac{Pr(T_{j(i)} = 1|X_j = x) \times Pr[S_j = 1]}{Pr(S_{j(i)} = 1|X_j = x) \times Pr[T_j = 1]}$$

Such an estimator is exactly equivalent to the matching estimator proposed by Imai and Kim (2019, eq. 14) setting $\hat{w}^t_{j(i)} = 1$ and restricting the matching set to switchers. Moreover, the weights are derived using the proposition proved by D. L. Miller et al. (2019), which establish that "appropriate group-level weight for [fixed effects] is proportional to the ratio of two propensity scores: (i) the propensity to be in the target population (e.g. program participants) and (ii) the propensity to be in the switcher population".[5] Of course, as previously stated in assumption (*v*), the functional form of the propensity score is crucial to achieve unbiasedness. Hence, even when standard literature states that linear probability models are enough, D. L. Miller et al. (2019) conclude that the functional form of the propensity scores does matter because; (i) in the family fixed effects design the identifying dimension (children per family) is very short, where most consistency results rely on at least four observations (ii) Various specifications handle variation differently, adding uncertainty to the estimates and in some cases even dropping observations. Although there is no clear informed consensus of which functional form to use, I mimic the *R* implementation of Imai and Kim (2019) (`package:wfe::pwfe`) to estimate the propensity scores using a bayesian logistic regression with a weakly informative distribution as recommended in Gelman et al. (2008). Such a model is robust to complete separation (high sparsity) and —in Gelman et al. (2008) words— "general enough to be used as a default in routine applied work".[6] Given that the frequentist and bayesian logistic regressions were qualitatively similar, I only report the ITT based on the bayesian method.

Validity of the Assumptions

Last but not least, I will discuss the theoretical compliance of this particular application to the both identification and statistical assumptions to estimate the ATE. In terms of the identification assumption ($*$) [*Selection into preschool among members of the same family is uncorrelated with the unobservable determinants of outcomes*], we have that: (i) As argued before, the preschool establishment rule was independent of any systematic family decision, rendering the family fixed effects design plausible. (ii) Even in the case parents were not equally altruistic among siblings (see Altonji et al., 1997 and Chang and Luo,

4 This assumption can be relaxed using the so-called Doubly-Robust estimators. Yet, it is left for future works.

5 As standard fixed effect estimations to recover treatment effects use $w_{FE} = {}^1/_{n_j} \sum_{i \in j}(D_i - {}^{\sum_{i \in j} \mathbf{1}(D_i=1)}/_{n_j})^2$, where n_j is the comparison group size; rather than population shares, $\hat{w}^t_{j(i)}$ is corrected by multiplying w_{FE}^{-1} (Gibbons et al., 2019).

6 In short, Gelman et al. (2008) "implement a procedure to fit generalized linear models in R with the Student-t prior distribution by incorporating an approximate EM algorithm into the usual iteratively weighted least squares", showing good fit in various applied applications ranging from demography to biology and bio-statistics. The implementation is done using `arm::bayesglm.fit(family=binomial(link="logit"))`.

2015 for extended discussion and models), after controlling for birth order effects and other family and individual characteristics, their willingness to send children to preschool —given that it is available— is plausibly independent of other factors taking into account that; First, even though there is evidence showing that in developing countries, parents indeed optimize sending their children to school and their allowance allocation based on their cognitive abilities (Akresh et al., 2012), the program's expansion would have reduced the costs to send the children to preschool, leveraging the opportunity costs of sending one extra child and thus ameliorating the self-selection concern.[7] Second, that parents' information about their children seems to be vague and not a good predictor of children's cognitive and non-cognitive abilities and nutrition —especially at early ages—, overriding a skill-optimization behavior from parents to children (S. A. Miller, 1986; Robinson & Sutin, 2016). (iii) To eliminate other possible confounders and capture only the effect related to the preschool expansion, I exclude from the switcher families those cases where the older sibling went to preschool, but the younger did not, and if having more than two siblings, those families that changed status more than once (e.g., if the older and younger siblings did not go to preschool, while the middle one did). That allows me to clean the estimated effect from unobservable variation related to unexpected economic shocks, favoritism, and other kinds of asymmetrical relations between parents and their children.

On the other hand, regarding statistical assumptions; About the first [*No spillover effects*], the very nature of education, framed in a human capital accumulation model, suggests that education spillovers across families are not common, and if they were widespread, their local nature would only support the stratification of workers' ability among neighborhoods —phenomenon called ghetto formation or poverty traps—. However, the literature suggests that the causality arrow goes from parental socioeconomic conditions to education of the children, meaning that clustering stem due to socioeconomic factors, rather than educational achievement (Durlauf, 1994). That, together with the fact that the existence of a preschool influences an entire area, makes the sole effect of one child attending preschool less likely to impact asymmetrically other neighboring children after controlling for family-level characteristics. The case of within-families spillovers is much more complicated, not only because data are scarce: recent literature shows that although there are significant education spillovers from older to younger siblings (see Black et al., 2021; Dahl et al., 2020 and Qureshi, 2018 for interesting examples), the effect is heavily dependent on the families' socioeconomic status, and sometimes even negative (Karbownik & Özek, 2019; Nicoletti & Rabe, 2019). According to Karbownik and Özek (2019), the differential net effects could be explained by two mechanisms: "mentoring/tutoring when the older sibling is positively affected by school-entry policies and parental reinforcement in more affluent households when either the younger or the older sibling is positively affected by school-entry policies". Therefore, it would imply that, if there indeed is a causal link between the educational performance of siblings, the effect of education would be underestimated, and interventions that improve the outcomes of one sibling could have larger benefits than those typically considered. Moreover, in the present setting, I am able to analyze preschool education at a national level, which would attenuate the effects of local education spillovers when differences among regions are evenly distributed.

Regarding the second [*There exists no unobserved group- or individual-varying confounders*] and third [*Conditional on observables, the treatment effect is independent of a unit's switching or*

[7] Indeed, Akresh et al. (2012) show the existence of a rivalry effect on attendance to primary school, which in the light of the universality of the Peruvian primary school, systematic optimization in preschool would seem implausible. However, in the Peruvian context such an effect could still have significant consequences later in life (e.g., high school and college).

target status], despite it is not possible to strictly rule out Roy-type "selection into treatment within groups" or an unobserved quality that increases the effectiveness of treatment, given that I cannot observe pre-treatment individual variables, as verified in chapter 4, preschool attendance appears to be uncorrelated with most family-varying observable characteristics of children when analyzing the switcher sample. However, individual covariates, which are not taken into account in the estimation of the propensity score, do vary systematically. In addition, if the program intended universalizing preschool, then, in goal terms, selection into treatment would not be plausible either. In case of the fourth [*Finite variance*], the methodology of the standardized tests assure that the variance of the scores is finite. It, of course, would not be enough if the measurement properties of the tests were not appropriate. However, as discussed in chapter 2, the tests are designed to measure the learning outcomes of the children in a theoretically compelling way, sufficient condition to satisfy the assumption.

Finally, albeit the last two assumptions cannot be formally proved without knowing the exact Data Generating Process [*Propensity scores are correctly defined*, and *There is a positive probability of being a switcher for each combination of group-level variables in the target group*], Sant'Anna and Song (2019) provide a specification test, which compares conditional cumulative distribution functions (CDFs). More specifically, they "derive a restriction between the propensity score CDFs among treated and control groups that gives information on overlapping, show that such a restriction is equivalent to a particular infinite number of unconditional moment conditions, and [base the] tests upon it". In that regard, the tests: (i) exploit the dimension-reduction coming from the derived restriction between propensity score CDFs, and (ii) conform an orthogonal projection onto the tangent space of nuisance parameters to add robustness for specification uncertainty. The null hypothesis of the test for this application is:

$$H_0 : \mathbb{E}[(D - \Lambda(X'\theta_0))\mathbf{1}\{\Lambda(X'\theta_0) \leq u\}] = 0 \text{ a.s. for some } \theta_0 \in \Theta \text{ and for all } u \in \Pi$$

where $\Lambda(.)$ is the logistic link function, $\Theta \subset \mathbb{R}^k$, and $\Pi = [0,1]$. Yet, reporting the test is not intended to be used as a model selection procedure but rather to point the modeling uncertainty of the ATE. In the end, only the systematic inclusion of relevant variables identified across different analyses and literature strands can get the current analysis closer to the best possible under ontological uncertainty.

Table 3.1 Variables and Sources of Data.

VARIABLE	LEVEL	TYPE	SOURCE
Test Score	*individual*	*numerical*	ECE
Female	*individual*	*indicator*	SIAGIE
Cohort	*individual*	*categorical*	SIAGIE
First-born	*individual*	*indicator*	SISFOH
Year with Preschool Eligibility	*individual*	*numerical*	SIAGIE/ECE
Birth-day Cut-off	*individual*	*indicator*	SIAGIE
Family ID	*family*	*index*	SISFOH
Socio-Economic-Status Index	*family*	*indicator*	SISFOH/ECE
Rurality	*family*	*indicator*	ECE
Household Size	*family*	*numerical*	SISFOH
Non-spanish Speaking	*family*	*indicator*	SISFOH/ECE
Participation in Social Programs	*family*	*indicator*	SISFOH
Access to formal Healthcare	*family*	*indicator*	SISFOH
Uniparental	*family*	*indicator*	SISFOH
Public School	*individual*	*indicator*	ECE
Single-Teacher School	*individual*	*indicator*	ECE

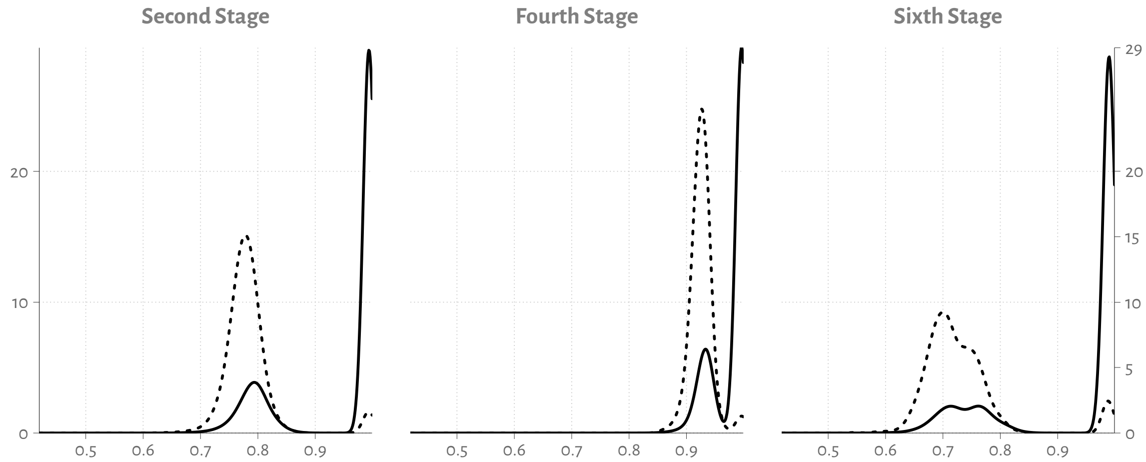
CHAPTER 4
Results and Discussion

Comparing between non-eligible and eligible groups by differentiating their switching status stresses the appropriateness of the re-weighting-on-observables method proposed by D. L. Miller et al. (2019) since it is evident that the switcher and non-switcher families differ significantly in terms of observable characteristics between eligible and non-eligible groups, which implies that the commonly recovered ATE (ITT in this case) could not be directly generalized into the broader population. However, even when for some of the variables across the samples there are no differences between eligible and non-eligible groups in the switcher group, the differences are mainly observed in individual-level covariates, a phenomenon which is itself generated by the identification framework. Such a characteristic does not pose too severe consequences if one could establish that there were no differences before the assignment. However, in this case, the lack of pre-treatment individual sibling data severely threatens the possible estimated effect. Instead, that presupposes an important limitation of the family fixed effects framework, since adding individual-varying covariates on the propensity score (weight) estimation breaks the identification of the treatment within families and, hence, requires some extensions which were not explored in the current analysis. Recalling chapter 3, all individual covariates were left out of the estimation of the propensity score and were instead added later in the weighted fixed effect regression to account for mean differences among the groups.

Furthermore, as seen in fig. 4.2 there is a clear lack of power derived from the small number of observations, in the case of the fourth and sixth stage samples. Indeed, the great heterogeneity across switcher and not-switcher samples also harmed the common support assumption, leading to complications in the estimation of the weights, as the propensity scores are not sufficiently apart from 0 and 1. Such a difficulty also precluded the calculation of the specification test proposed by Sant'Anna and Song (2019).

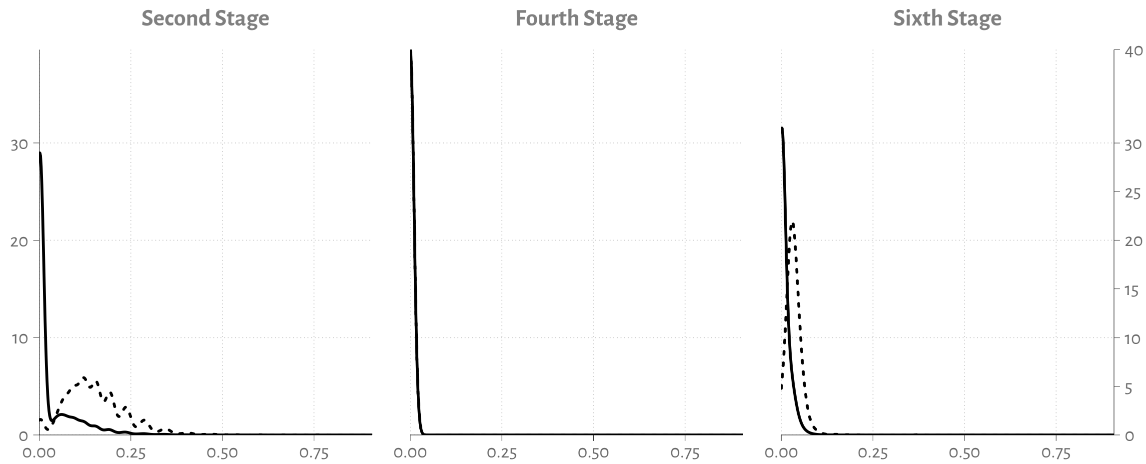
As stated before, a statistical test is not enough to override a set of regressors in the light of an uncertain data generating process. In that sense, another interesting point to highlight is that all employed covariates are significantly different between eligible and non-eligible groups in the whole sample for all three analyzed tests, confirming why they are theoretically and empirically relevant. This, in turn, suggests that using them to estimate the propensity score is important to achieve balance in the matching and recover the ATE for the broader population because, as stated before, the functional form of the propensity score is fundamental to recover the unbiased effect. Nonetheless, since most of them are categorical or dummy variables, it could be that the information conveyed is not enough to estimate a model as the one considered here, explaining the lack of support along the propensity score distribution and its excessive concentration. Apropos, although I had access to national-wide databases, containing millions of observations, the estimation of the treatment under this framework, and with such intensive use of indicators, requires huge amounts of data, as the identification rests on very special conditions. It could further suggest that employing a differences-in-differences estimator could be more appropriate, as the identification conditions are somewhat weaker. However, and although I already discussed why a traditional difference-in-differences is not suitable for this particular case, exploring variations on the identification of the family-fixed effect while controlling simultaneously for time variation seems like a promising avenue for this analysis. At the time of this study, most of the developed two-way fixed effects solutions that take into account the inherent bias of the fixed effect estimators were defined on an individual level, with no simultaneous within-group debiasing.

Figure 4.1 P-Score Overlap:
Estimated propensity to be eligible to preschool $Pr(T_{j(i)} = 1|X_j = x) \times Pr[S_j = 1]$.



This figure shows kernel density plots (bandwidth = 0.01) of the predicted probability of having a preschool available, for switchers [dotted line] and non-switchers [solid line] that had a preschool available.

Figure 4.2 P-Score Overlap:
Estimated propensity to be a switcher family $Pr(S_{j(i)} = 1|X_j = x) \times Pr[T_j = 1]$.



This figure shows kernel density plots (bandwidth = 0.01) of the predicted probability of belonging to a switcher group, for switchers [dotted line] and non-switchers [solid line] that were eligible to attend preschool.

All in all, given the above reasons, I do not feel it appropriate to interpret the obtained regression coefficients as the ITT, nor further analyzing them. However, as seen in the table 4.4, the estimates for the effect on second stage outcomes seem to be much more accurate (in terms of standard errors) than the unweighted "vanilla" fixed effects in table 4.5. Concluding, I would like to address the following; (i) The trade-off between internal and external validity seems more valid than ever, as traditional methods would be evidently biased with respect to the population estimate, while innovative approaches such as this one could lack power to alleviate the trade-off. (ii) The natural step after this analysis would be considering the extended family-fixed effects estimator, which, given space constraints, was not reproduced. However, it is also reasonable to acknowledge that it could be prone to the same weaknesses exposed here. (iii) Finally, further explorations should be carried out concerning the huge differences between eligible and non-eligible groups, as well as switcher and non-switcher families. A fact that, being more than a statistical impasse, could also shed light on the dynamics of education, human capital accumulation, and public policy in one of the most unequal regions in the world: Latin America.

Table 4.1 Selected Family and Individual Characteristics, by Preschool Availability and Switcher Status —Second Stage Test Sample.

	Whole Sample			Non-Switchers			Switchers		
	Non-Elegible $N=118132$	Elegible $N=1964701$	*p-value*	Non-Elegible $N=85916$	Elegible $N=1933661$	*p-value*	Non-Elegible $N=32216$	Elegible $N=31040$	*p-value*
Female	56451 (47.8%)	932919 (47.5%)	0.044	40905 (47.6%)	917788 (47.5%)	0.401	15546 (48.3%)	15131 (48.7%)	0.219
Cohort:			0.000			0.000			0.000
$[2002, 2005)$	9009 (7.63%)	43860 (2.23%)		6173 (7.18%)	43793 (2.26%)		2836 (8.80%)	67 (0.22%)	
$[2005, 2010)$	81572 (69.1%)	795974 (40.5%)		56501 (65.8%)	792242 (41.0%)		25071 (77.8%)	3732 (12.0%)	
$[2010, 2013)$	23093 (19.5%)	679743 (34.6%)		18981 (22.1%)	664513 (34.4%)		4112 (12.8%)	15230 (49.1%)	
$[2013, 2017)$	4458 (3.77%)	445124 (22.7%)		4261 (4.96%)	433113 (22.4%)		197 (0.61%)	12011 (38.7%)	
First born = 1	33938 (28.8%)	597604 (30.4%)	<0.001	23979 (28.0%)	597233 (30.9%)	<0.001	9959 (30.9%)	371 (1.20%)	0.000
Born before March 31st = 1	34510 (29.2%)	514003 (26.2%)	<0.001	24993 (29.1%)	507281 (26.2%)	<0.001	9517 (29.5%)	6722 (21.7%)	<0.001
Public School = 1	115885 (98.1%)	1509442 (76.8%)	0.000	84033 (97.8%)	1478782 (76.5%)	0.000	31852 (98.9%)	30660 (98.8%)	0.288
Single-Teacher School = 1	98656 (83.5%)	311267 (15.8%)	0.000	71695 (83.4%)	290869 (15.0%)	0.000	26961 (83.7%)	20398 (65.7%)	0.000
Socio-Economic Status Index	0.81 [-0.16;1.75]	0.59 [-1.14;2.23]	<0.001	0.81 [-0.16;1.77]	0.58 [-1.17;2.24]	<0.001	0.81 [-0.16;1.71]	0.80 [-0.16;1.69]	0.070
Rural = 1	110624 (93.6%)	418012 (21.3%)	0.000	79633 (92.7%)	388296 (20.1%)	0.000	30991 (96.2%)	29716 (95.7%)	0.003
Household Size	6.00 [4.00;7.00]	5.00 [4.00;6.00]	0.000	5.00 [4.00;7.00]	5.00 [4.00;6.00]	0.000	6.00 [5.00;8.00]	6.00 [5.00;8.00]	0.164
No Spanish at Home = 1	5796 (4.91%)	127282 (6.48%)	<0.001	4869 (5.67%)	126370 (6.54%)	<0.001	927 (2.88%)	912 (2.94%)	0.667
Enrolled in Social Programs = 1	63406 (53.7%)	602324 (30.7%)	0.000	40950 (47.7%)	580640 (30.0%)	0.000	22456 (69.7%)	21684 (69.9%)	0.680
Access to Healthcare = 1	99089 (84.0%)	1407823 (71.8%)	0.000	70632 (82.4%)	1380390 (71.5%)	0.000	28457 (88.4%)	27433 (88.5%)	0.844
Single-Parent Family = 1	13963 (11.8%)	299005 (15.2%)	<0.001	11532 (13.4%)	296659 (15.3%)	<0.001	2431 (7.55%)	2346 (7.56%)	0.966

Numerical variables were compared using Wilcoxon's Ranked Tests. For all the variables and groups the effective number of observations is reported.

P-values are calculated within each group, and are adjusted for multiple comparisons. Brackets report s.e. at the 0.95 confidence level.

Table 4.2 Selected Family and Individual Characteristics, by Preschool Availability and Switcher Status —Fourth Stage Test Sample.

	Whole Sample			Non-Switchers			Switchers		
	Non-Elegible N =9005	Elegible N =512136	*p-value*	Non-Elegible N =8245	Elegible N =511363	*p-value*	Non-Elegible N =760	Elegible N =773	*p-value*
Female	4335 (48.1%)	247056 (48.2%)	0.858	3971 (48.2%)	246699 (48.2%)	0.893	364 (47.9%)	357 (46.2%)	0.535
Cohort:			0.000			0.000			<0.001
$[2005, 2010)$	521 (5.79%)	2264 (0.44%)		406 (4.92%)	2259 (0.44%)		115 (15.1%)	5 (0.65%)	
$[2010, 2013)$	6322 (70.2%)	246957 (48.2%)		5697 (69.1%)	246769 (48.3%)		625 (82.2%)	188 (24.3%)	
$[2013, 2017)$	2162 (24.0%)	262914 (51.3%)		2142 (26.0%)	262334 (51.3%)		20 (2.63%)	580 (75.0%)	
First born = 1	1939 (21.6%)	139664 (27.3%)	<0.001	1795 (21.8%)	139649 (27.3%)	<0.001	144 (19.0%)	15 (1.94%)	<0.001
Born before March 31st = 1	2856 (31.7%)	137344 (26.8%)	<0.001	2642 (32.0%)	137180 (26.8%)	<0.001	214 (28.2%)	164 (21.2%)	0.002
Public School = 1	8624 (95.8%)	400043 (78.1%)	0.000	7871 (95.5%)	399274 (78.1%)	<0.001	753 (99.1%)	769 (99.5%)	0.526
Single-Teacher School = 1	6010 (66.7%)	75613 (14.8%)	0.000	5433 (65.9%)	75074 (14.7%)	0.000	577 (75.9%)	539 (69.7%)	0.008
Socio-Economic Status Index	1.09 [0.02;1.99]	0.60 [-1.07;2.22]	<0.001	1.07 [0.01;1.99]	0.60 [-1.07;2.22]	<0.001	1.17 [0.20;2.04]	1.18 [0.24;2.04]	0.793
Rural = 1	7684 (85.3%)	102291 (20.0%)	0.000	6953 (84.3%)	101547 (19.9%)	0.000	731 (96.2%)	744 (96.2%)	1.000
Household Size	5.00 [4.00;7.00]	5.00 [4.00;6.00]	<0.001	5.00 [4.00;7.00]	5.00 [4.00;6.00]	<0.001	7.00 [5.00;8.00]	7.00 [5.00;8.00]	0.639
No Spanish at Home = 1	1123 (12.5%)	76547 (14.9%)	<0.001	1102 (13.4%)	76525 (15.0%)	<0.001	21 (2.76%)	22 (2.85%)	1.000
Enrolled in Social Programs = 1	5232 (58.1%)	188067 (36.7%)	0.000	4659 (56.5%)	187486 (36.7%)	<0.001	573 (75.4%)	581 (75.2%)	0.963
Access to Healthcare = 1	7415 (82.5%)	373802 (73.1%)	<0.001	6741 (81.9%)	373114 (73.1%)	<0.001	674 (88.8%)	688 (89.1%)	0.907
Single-Parent Family = 1	908 (10.1%)	72387 (14.1%)	<0.001	868 (10.5%)	72347 (14.1%)	<0.001	40 (5.26%)	40 (5.17%)	1.000

Numerical variables were compared using Wilcoxon's Ranked Tests. For all the variables and groups the effective number of observations is reported.

P-values are calculated within each group, and are adjusted for multiple comparisons. Brackets report S.E. at the 0.95 confidence level.

Table 4.3 Selected Family and Individual Characteristics, by Preschool Availability and Switcher Status —Sixth Stage Test Sample.

	Whole Sample			Non-Switchers			Switchers		
	Non-Elegible N =60895	Elegible N =771862	*p-value*	Non-Elegible N =57703	Elegible N =768778	*p-value*	Non-Elegible N =3192	Elegible N =3084	*p-value*
Female	28744 (47.2%)	373490 (48.4%)	<0.001	27262 (47.2%)	371963 (48.4%)	<0.001	1482 (46.4%)	1527 (49.5%)	0.016
Cohort:			.			.			0.000
$[2002, 2005)$	1386 (2.28%)	4418 (0.57%)		1205 (2.09%)	4414 (0.57%)		181 (5.67%)	4 (0.13%)	
$[2005, 2010)$	50673 (83.2%)	572884 (74.2%)		47689 (82.6%)	571956 (74.4%)		2984 (93.5%)	928 (30.1%)	
$[2010, 2013)$	8834 (14.5%)	194550 (25.2%)		8808 (15.3%)	192399 (25.0%)		26 (0.81%)	2151 (69.7%)	
First born = 1	17953 (29.5%)	254454 (33.0%)	<0.001	16972 (29.4%)	254415 (33.1%)	<0.001	981 (30.7%)	39 (1.27%)	<0.001
Born before March 31st = 1	17328 (28.5%)	195437 (25.3%)	<0.001	16374 (28.4%)	194925 (25.4%)	<0.001	954 (29.9%)	512 (16.6%)	<0.001
Public School = 1	58264 (95.7%)	611632 (79.2%)	0.000	55175 (95.6%)	608600 (79.2%)	0.000	3089 (96.8%)	3032 (98.3%)	<0.001
Single-Teacher School = 1	.	.	.	.	.	.	.	.	.
Socio-Economic Status Index	0.70 [-0.16;1.64]	0.53 [-1.14;2.11]	<0.001	0.71 [-0.16;1.64]	0.53 [-1.14;2.12]	<0.001	0.65 [-0.16;1.57]	0.65 [-0.16;1.57]	0.976
Rural = 1	55876 (91.8%)	156656 (20.3%)	0.000	52901 (91.7%)	153780 (20.0%)	0.000	2975 (93.2%)	2876 (93.3%)	0.972
Household Size	6.00 [4.00;7.00]	5.00 [4.00;6.00]	0.000	6.00 [4.00;7.00]	5.00 [4.00;6.00]	0.000	6.00 [5.00;8.00]	6.00 [5.00;8.00]	0.421
No Spanish at Home = 1	5700 (9.36%)	114317 (14.8%)	<0.001	5581 (9.67%)	114201 (14.9%)	<0.001	119 (3.73%)	116 (3.76%)	0.998
Enrolled in Social Programs = 1	33531 (55.1%)	219831 (28.5%)	0.000	31613 (54.8%)	217987 (28.4%)	0.000	1918 (60.1%)	1844 (59.8%)	0.831
Access to Healthcare = 1	52027 (85.6%)	556846 (72.2%)	0.000	49183 (85.4%)	554111 (72.2%)	0.000	2844 (89.3%)	2735 (88.9%)	0.633
Single-Parent Family = 1	6726 (11%)	120255 (15%)	<0.001	6480 (11.2%)	120017 (15.6%)	<0.001	246 (7.71%)	238 (7.72%)	1.000

Numerical variables were compared using Wilcoxon's Ranked Tests. For all the variables and groups the effective number of observations is reported.

P-values are calculated within each group, and are adjusted for multiple comparisons. Brackets report S.E. at the 0.95 confidence level.

"." refers to completely missing variables. Cohort $(2013, 2017]$ was dropped for this sample.

Table 4.4 Weighted Fixed Effects Estimation à la D. L. Miller et al. (2019) — Second Stage Test [ECE].

	Mathematics	Language
Preschool (ATE)	−0.538	−0.653
	[−1.405, 0.330]	[−1.214, −0.092]
	S.E. = 0.442	S.E. = 0.286
	p-value = 0.224	*p-value* = 0.022
Female = 1	−0.122	−0.044
	[−0.274, 0.030]	[−0.154, 0.067]
	S.E. = 0.077	S.E. = 0.057
	p-value = 0.116	*p-value* = 0.441
Cohort $[2005, 2010)$	−0.045	0.346
	[−0.867, 0.776]	[−0.153, 0.846]
	S.E. = 0.419	S.E. = 0.255
	p-value = 0.914	*p-value* = 0.174
Cohort $[2010, 2013)$	0.285	0.692
	[−0.528, 1.098]	[0.197, 1.187]
	S.E. = 0.415	S.E. = 0.253
	p-value = 0.492	*p-value* = 0.006
Cohort $[2013, 2017)$	0.802	0.938
	[−0.017, 1.622]	[0.441, 1.434]
	S.E. = 0.418	S.E. = 0.253
	p-value = 0.055	*p-value* = 0.000
First born = 1	−0.064	−0.258
	[−0.662, 0.535]	[−0.792, 0.276]
	S.E. = 0.305	S.E. = 0.272
	p-value = 0.835	*p-value* = 0.344
Born before March 31st = 1	−0.019	0.036
	[−0.206, 0.168]	[−0.100, 0.173]
	S.E. = 0.095	S.E. = 0.070
	p-value = 0.845	*p-value* = 0.604
Public School = 1	0.659	0.430
	[0.327, 0.992]	[0.123, 0.738]
	S.E. = 0.170	S.E. = 0.157
	p-value = 0.000	*p-value* = 0.006
Single-Teacher School = 1	−0.195	−0.089
	[−0.350, −0.040]	[−0.212, 0.034]
	S.E. = 0.079	S.E. = 0.063
	p-value = 0.014	*p-value* = 0.158
Num.Obs.	60 856	60 811
Std.Errors	Family-level	

In brackets correspond to the confidence intervals at the 0.95 confidence level. All variables different from Preschool (ATE) are interaction between Preschool and each corresponding row. Outcomes are standardized with respect to the ECE Control Sample in 2017, that is, subtracting a mean = 500 and a S.D. = 100. Coefficients can then be interpreted as S.D.

Table 4.5 Unweighted Fixed Effects Estimation — Second Stage Test [ECE].

	Mathematics	Language
Preschool (ATE)	−0.043	−0.069
	[−0.474, 0.389]	[−0.370, 0.232]
	s.e.= 0.220	s.e.= 0.154
	p-value= 0.847	*p-value*= 0.654
Female $= 1$	−0.035	0.013
	[−0.072, 0.003]	[−0.014, 0.040]
	s.e.= 0.019	s.e.= 0.014
	p-value= 0.072	*p-value*= 0.339
Cohort $[2005, 2010)$	−0.223	0.126
	[−0.610, 0.163]	[−0.139, 0.392]
	s.e.= 0.197	s.e.= 0.136
	p-value= 0.258	*p-value*= 0.351
Cohort $[2010, 2013)$	−0.007	0.411
	[−0.393, 0.379]	[0.145, 0.676]
	s.e.= 0.197	s.e.= 0.136
	p-value= 0.972	*p-value*= 0.002
Cohort $[2013, 2017)$	0.410	0.642
	[0.023, 0.797]	[0.376, 0.907]
	s.e.= 0.197	s.e.= 0.136
	p-value= 0.038	*p-value*= 0.000
First born $= 1$	−0.048	−0.159
	[−0.225, 0.130]	[−0.305, −0.012]
	s.e.= 0.091	s.e.= 0.075
	p-value= 0.600	*p-value*= 0.034
Born before March 31st $= 1$	−0.009	0.019
	[−0.054, 0.037]	[−0.013, 0.052]
	s.e.= 0.023	s.e.= 0.016
	p-value= 0.701	*p-value*= 0.237
Public School $= 1$	0.376	0.140
	[0.183, 0.569]	[−0.004, 0.283]
	s.e.= 0.098	s.e.= 0.073
	p-value= 0.000	*p-value*= 0.056
Single-Teacher School $= 1$	−0.224	−0.175
	[−0.268, −0.180]	[−0.207, −0.143]
	s.e.= 0.022	s.e.= 0.016
	p-value= 0.000	*p-value*= 0.000
Num.Obs.	60 856	60 811
Std.Errors	Family-level	

In brackets correspond to the confidence intervals at the 0.95 confidence level. All variables different from Preschool (ATE) are interaction between Preschool and each corresponding row. Outcomes are standardized with respect to the ECE Control Sample in 2017, that is, subtracting a mean $= 500$ and a s.d. $= 100$. Coefficients can then be interpreted as s.d.

Table 4.6 Unweighted Fixed Effects Estimation — Fourth Stage Test [ECE].

	Mathematics	Language
Preschool (ATE)	0.727	0.867
	[−0.325, 1.779]	[−0.256, 1.991]
	S.E. = 0.536	S.E. = 0.572
	p-value = 0.175	*p-value* = 0.130
Female = 1	−0.198	−0.067
	[−0.389, −0.007]	[−0.251, 0.116]
	S.E. = 0.097	S.E. = 0.093
	p-value = 0.042	*p-value* = 0.470
Cohort $[2010, 2013)$	−0.533	−0.359
	[−1.286, 0.219]	[−1.347, 0.630]
	S.E. = 0.383	S.E. = 0.503
	p-value = 0.164	*p-value* = 0.476
Cohort $[2013, 2017)$	−0.492	−0.450
	[−1.232, 0.248]	[−1.424, 0.524]
	S.E. = 0.377	S.E. = 0.496
	p-value = 0.192	*p-value* = 0.365
First born = 1	0.032	0.102
	[−0.501, 0.566]	[−0.587, 0.791]
	S.E. = 0.272	S.E. = 0.351
	p-value = 0.906	*p-value* = 0.772
Born before March 31st = 1	−0.052	−0.157
	[−0.281, 0.177]	[−0.378, 0.065]
	S.E. = 0.117	S.E. = 0.113
	p-value = 0.657	*p-value* = 0.165
Public School = 1	−0.015	−0.319
	[−0.786, 0.756]	[−0.909, 0.272]
	S.E. = 0.393	S.E. = 0.301
	p-value = 0.970	*p-value* = 0.290
Single-Teacher School = 1	−0.212	−0.133
	[−0.431, 0.007]	[−0.352, 0.086]
	S.E. = 0.111	S.E. = 0.112
	p-value = 0.058	*p-value* = 0.234
Num.Obs.	1457	1458
Std.Errors	Family-level	

In brackets correspond to the confidence intervals at the 0.95 confidence level. All variables different from Preschool (ATE) are interaction between Preschool and each corresponding row. Outcomes are standardized with respect to the ECE Control Sample in 2017, that is, subtracting a mean $= 500$ and a S.D. $= 100$. Coefficients can then be interpreted as S.D.

Table 4.7 Unweighted Fixed Effects Estimation — Sixth Stage Test [ECE].

	Mathematics	Language
Preschool (ATE)	0.109	−0.729
	[−0.488, 0.705]	[−1.304, −0.154]
	S.E. = 0.304	S.E. = 0.293
	p-value = 0.721	*p-value* = 0.013
Female = 1	−0.118	−0.009
	[−0.189, −0.048]	[−0.067, 0.050]
	S.E. = 0.036	S.E. = 0.030
	p-value = 0.001	*p-value* = 0.768
Cohort $[2005, 2010)$	−0.180	0.629
	[−0.638, 0.277]	[0.118, 1.141]
	S.E. = 0.233	S.E. = 0.261
	p-value = 0.440	*p-value* = 0.016
Cohort $[2010, 2013)$	−0.210	0.638
	[−0.669, 0.249]	[0.127, 1.150]
	S.E. = 0.234	S.E. = 0.261
	p-value = 0.370	*p-value* = 0.014
First born = 1	−0.130	0.005
	[−0.471, 0.211]	[−0.241, 0.252]
	S.E. = 0.174	S.E. = 0.126
	p-value = 0.455	*p-value* = 0.966
Born before March 31st = 1	−0.026	0.029
	[−0.154, 0.103]	[−0.083, 0.142]
	S.E. = 0.065	S.E. = 0.057
	p-value = 0.697	*p-value* = 0.607
Public School = 1	0.167	0.216
	[−0.216, 0.549]	[−0.043, 0.475]
	S.E. = 0.195	S.E. = 0.132
	p-value = 0.393	*p-value* = 0.103
Num.Obs.	6018	6020
Std.Errors	Family-level	

In brackets correspond to the confidence intervals at the 0.95 confidence level. All variables different from Preschool (ATE) are interaction between Preschool and each corresponding row. Outcomes are standardized with respect to the ECE Control Sample in 2017, that is, subtracting a mean $= 500$ and a S.D. $= 100$. Coefficients can then be interpreted as S.D. Given the lack of observations for the cohort $[2013, 2017)$, it was left out from the estimation.

References

Akresh, R., Bagby, E., De Walque, D., & Kazianga, H. (2012). Child ability and household human capital investment decisions in burkina faso. *Economic Development and Cultural Change*, *61*(1), 157–186.

Altonji, J. G., Hayashi, F., & Kotlikoff, L. J. (1997). Parental altruism and inter vivos transfers: Theory and evidence. *Journal of political economy*, *105*(6), 1121–1166.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC, USA.

Ames, P. (2016). La educación básica peruana a inicios del siglo xxi : Posibilidades y desafíos. In M. Quero (Ed.), *El perú en los incicios del siglo xxi. cambios y continuidades desde las ciencias sociales* (pp. 29–46). Mexico City, Mexico, Universidad Nacional Autónoma de México.

Asmat Linares, M. M. (2020). *Identificación de centros funcionales y sus áreas de influencia para la identificación de acciones de organización territorial.* (Boletín No. 5). Colegio de Geógrafos del Perú. Lima, Perú.

Barclay, K. J., Lyngstad, T. H., & Conley, D. C. (2021). The production of inequalities within families and across generations: The intergenerational effects of birth order on educational attainment. *European Sociological Review*.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*(6045), 975–978.

Behrman, J. R., & Urzúa, S. S. (2013). *Economic perspectives on some important dimensions of early childhood development in developing countries* (P. R. Britto, P. L. Engle, & C. M. Super, Eds.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199922994.001.0001

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.). (2020). *Comparability of large-scale educational assessments: Issues and recommendations.* Washington DC, USA, National Academy of Education.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., Roth, J., & Simonsen, M. (2021). Sibling spillovers. *The Economic Journal*, *131*(633), 101–128.

Bofah, E. A.-t., & Hannula, M. S. (2017). Home resources as a measure of socio-economic status in ghana. *Large-scale Assessments in Education*, *5*(1), 1–15.

Bonal, X. (2007). On global absences: Reflections on the failings in the education and poverty relationship in latin america. *International Journal of Educational Development*, *27*(1), 86–100.

Britto, P. R., Engle, P. L., & Super, C. M. (Eds.). (2013). *Handbook of early childhood development research and its impact on global policy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199922994.001.0001

Britto, P. R., Yoshikawa, H., Ravens, J. V., Ponguta, L. A., Oh, S. S., Dimaya, R., & Seder, R. C. (2013). Understanding governance of early childhood development and education systems and services in low-income countries, 66.

Burkett, K., Pickler, R., Bowers, K., & Folker, A. E. (2020). Disparities affect developmental risk for head start preschoolers. *Journal of Pediatric Nursing*, *54*, 86–92.

Carswell, A. T. (2015). Neighborhood organizations. In F. Wherry & J. Schor (Eds.), *The SAGE encyclopedia of economics and society* (pp. 1168–1169). SAGE Publications, Inc. https://doi.org/10.4135/9781452206905

Center for Global Education Monitoring. (2014). *The latin-american laboratory for assessment of the quality of education: Measuring and comparing educational quality in latin america* (Assessment GEM Series: No 3).

Chang, Y.-M., & Luo, Z. (2015). Endogenous division rules as a family constitution: Strategic altruistic transfers and sibling competition. *Journal of Population Economics*, *28*(1), 173–194.

Chernozhukov, V., Fernández-Val, I., Hahn, J., & Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, *81*(2), 535–580.

Contreras, D., Delgadillo, J., & Riveros, G. (2019). Is home overcrowding a significant factor in children's academic performance? evidence from latin america. *International Journal of Educational Development*, *67*, 1–17.

Córdova, A. (2009). Methodological note: Measuring relative wealth using household asset indicators. *AmericasBarometer Insights*, *6*, 1–9.

Dahl, G. B., Rooth, D.-O., & Stenberg, A. (2020). *Family spillovers in field of study*. National Bureau of Economic Research.

De Haan, M., Plug, E., & Rosero, J. (2014). Birth order and human capital development evidence from ecuador. *Journal of Human Resources*, *49*(2), 359–392.

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

Dirección Nacional de Educación Inicial y Primaria. (2003). *Plan estratégico de educación inicial*. Ministerio de Educación. Lima, Peru.

Duarte, J., Bos, M. S., & Moreno, M. (2010). Inequity in school achievement in latin america. *Multilevel Analysis of SERCE Results According to the Socioeconomic Status of Students. New York: IADB*.

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of economic perspectives*, *27*(2), 109–32.

Durlauf, S. N. (1994). Spillovers, stratification, and inequality. *European Economic Review*, *38*(3-4), 836–845.

Engle, P. L., Rao, N., & Petrovic, O. (2013). *Situational analysis of young children in a changing world* (P. R. Britto, P. L. Engle, & C. M. Super, Eds.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199922994.001.0001

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., Et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise., In *Kdd*.

Gao, Q.-L., Li, Q.-Q., Zhuang, Y., Yue, Y., Liu, Z.-Z., Li, S.-Q., & Sui, D. (2019). Urban commuting dynamics in response to public transit upgrades: A big data approach. *PloS one*, *14*(10), e0223650.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The annals of applied statistics*, *2*(4), 1360–1383.

General education bill [Law No. 28044]. (2004). *Lima, Peru*. https://bit.ly/3uFCnAA Approved by the Congress of the Republic, promulgated by the President.

Gibbons, C. E., Serrato, J. C. S., & Urbancic, M. B. (2019). Broken or fixed effects? *Journal of Econometric Methods*, *8*(1).

Gil, J., & Read, S. (2014). Patterns of sustainable mobility and the structure of modality in the randstad city-region. *A| Z ITU Journal of the Faculty of Architecture*, *11*(2), 231–254.

Gob.pe. (2021). *Sistema de focalización de hogares (sisfoh)*. Retrieved October 15, 2021, from https://www.gob.pe/437-sistema-de-focalizacion-de-hogares-sisfoh

González, J., & Gempp, R. (2021). Test comparability and measurement validity in educational assessment, In *Validity of educational assessments in chile and latin america*. Springer.

Gregosz, D., Perticará, M., & Loebel-Radefeldt, K. (Eds.). (2014). *Los Desafíos de Educación Preescolar, Básica y Media en América Latina*. Santiago, Chile, Konrad-Adenauer-Stiftung e.V.

Guadalupe, C. (2015). Aprendizajes desde la ece y sus perspectivas. *Revista TAREA*, (89), 9–15.

Guadalupe, C., Burga, A., Miranda, L., & Castillo, L. E. (2015). Brechas de equidad en la evaluación censal de estudiantes 2007-2014: Tres aproximaciones a su medición. *Persona*, (18), 47–68.

Guadalupe, C., León, J., Rodríguez, J. S., & Vargas, S. (2017). *Estado de la educación en el perú: Análisis y perspectivas de la educación básica*. Proyecto Fortalecimiento de la Gestión de la Educación en el Perú (FORGE). Lima, Perú, Group for the Analysis of Development (GRADE).

Guerrero, G., & Demarini, F. (2016). Atención y educación de la primera infancia en el Perú: avances y retos pendientes. In S. Cueto (Ed.), *Investigación para el desarrollo en el Perú: Once balances* (pp. 163–206). Lima, Perú, Group for the Analysis of Development (GRADE).

Imai, K., & Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science*, *63*(2), 467–490.

Instituto Nacional de Estadística e Informática. (2018). *Perú: Crecimiento y distribución de la población, 2017* (Primeros Resultados). Lima, Perú.

Instituto Nacional de Estadística e Informática. (2019). *Indicadores de educación por departamento 2008-2018*. Lima, Perú.

Instituto Nacional de Estadística e Informática. (2021). *Perú: Estado de la población en el año del bicentenario, 2021*. Lima, Perú.

Kabudula, C. W., Houle, B., Collinson, M. A., Kahn, K., Tollman, S., & Clark, S. (2017). Assessing changes in household socioeconomic status in rural south africa, 2001–2013: A distributional analysis using household asset indicators. *Social Indicators Research*, *133*(3), 1047–1073.

Kagan, S. L. (1990). Readiness 2000: Rethinking rhetoric and responsibility. *The Phi Delta Kappan*, *72*(4), 272–279.

Karbownik, K., & Özek, U. (2019). *Setting a good example? examining sibling spillovers in educational achievement using a regression discontinuity design*. National Bureau of Economic Research.

Lee, D., & McLanahan, S. (2015). Family structure transitions and child development: Instability, selection, and population heterogeneity. *American sociological review*, *80*(4), 738–763.

Lehti, H., Erola, J., & Karhula, A. (2019). The heterogeneous effects of parental unemployment on siblings' educational outcomes. *Research in Social Stratification and Mobility*, *64*, 100439.

León Mendoza, J. (2019). Efectos de la desnutrición en el rendimiento escolar de los niños en el Perú. *Pensamiento Crítico*, *24*(1), 79–102. https://doi.org/10.15381/pc.v24i1.16560

Llanos Zuloaga, M. (2016). Pioneering programs of non formal early childhood education in Peru and its impact on the overcoming of educational inequality of women. *Avances en Psicología*, *24*(1), 33–49.

Lu, C., Cuartas, J., Fink, G., McCoy, D., Liu, K., Li, Z., Daelmans, B., & Richter, L. (2020). Inequalities in early childhood care and development in low/middle-income countries: 2010–2018. *BMJ Global Health*, *5*(2), e002314.

Majerowicz Nieto, S. (2019). *Essays in education and development economics: The effect of a preschool expansion on early learning outcomes in peru* (Doctoral dissertation). Harvard University.

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205.

Miller, D. L., Shenhav, N., & Grosz, M. Z. (2019). *Selection into identification in fixed effects models, with application to head start*. National Bureau of Economic Research.

Miller, S. A. (1986). Parents' beliefs about their children's cognitive abilities. *Developmental Psychology*, *22*(2), 276.

Ministerial decree no. 529-2017-minedu. (2017). *Lima, Peru*. https://bit.ly/3DfSZCp

Ministerio de Educación. (2016). *Marco de fundamentación de las pruebas de la evaluación censal de estudiantes* (Series: Investigaciones). Lima, Perú.

Ministerio de Educación. (2017). *Perú ¿cómo vamos en educación?* Lima, Perú, Repositorio Institucional.

Ministerio de Educación. (2017). *Reporte técnico del estatus socioeconómico en la Evaluación Censal de Estudiantes 2016*. Lima, Perú.

Morales, M. (2020). *School-entry eligibility effects in developing countries* (Doctoral dissertation). Columbia University.

Navarro, N. P. Et al. (2018). *Ethnic diversity in peruvian schools: Disentangling peer and class composition effects*. Centre for the Study of African Economies, University of Oxford.

Neidhöfer, G., Serrano, J., & Gasparini, L. (2018). Educational inequality and intergenerational mobility in latin america: A new database. *Journal of Development Economics*, *134*, 329–349.

Nicoletti, C., & Rabe, B. (2019). Sibling spillover effects in school achievement. *Journal of Applied Econometrics*, *34*(4), 482–501.

Nores, M. (2020). The economics of early childhood interventions, In *The economics of education: A comprehensive overview*. Elsevier.

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world:(under) investing in the very young. *Economics of Education Review*, *29*(2), 271–282.

Oficina de Infraestructura Educativa. (2011). *Normas técnicas para el diseño de locales de educación básica regular: Nivel inicial*. Ministerio de Educación. Lima, Perú.

Ortiz Chávez, P., & Pastor Vargas, C. (2014). Los desafíos de la educación preescolar, básica y media en el Perú. In D. Gregosz, M. Perticará, & K. Loebel-Radefeldt (Eds.), *Los Desafíos de Educación Preescolar, Básica y Media en América Latina* (pp. 251–270). Santiago, Chile, Konrad-Adenauer-Stiftung e.V.

Pages, R., Lukes, D. J., Bailey, D. H., & Duncan, G. J. (2020). Elusive longer-run impacts of head start: Replications within and across cohorts. *Educational Evaluation and Policy Analysis*, *42*(4), 471–492.

Pineda Zumaran, J. (2016). Spatial data usage, spatial thinking and spatial knowledge generation: The case of planning practitioners in arequipa, peru. *Planning Practice & Research*, *31*(3), 270–291.

Prime, H., Plamondon, A., & Jenkins, J. M. (2017). Birth order and preschool children's cooperative abilities: A within-family analysis. *British Journal of Developmental Psychology*, *35*(3), 392–405.

Qureshi, J. A. (2018). Siblings, teachers, and spillovers on academic achievement. *Journal of Human Resources*, *53*(1), 272–297.

Richter, L. M., Cappa, C., Issa, G., Lu, C., Petrowski, N., & Naicker, S. N. (2020). Data for action on early childhood development. *The Lancet*, *396*(10265), 1784–1786.

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., Lu, C., Lucas, J. E., Perez-Escamilla, R., Dua, T., Et al. (2017). Investing in the foundation of sustainable development: Pathways to scale up for early childhood development. *The lancet*, *389*(10064), 103–118.

Rivas, M. E., Suarez-Aleman, A., & Serebrisky, T. (2019). *Stylized urban transportation facts in latin america and the caribbean* (Technical Note No. 1640). Inter-American Development Bank.

Robinson, E., & Sutin, A. R. (2016). Parental perception of weight status and weight gain across childhood. *Pediatrics*, *137*(5).

Rodríguez-Morata, C., Díaz, H., Ballesteros-Canovas, J. A., Rohrer, M., & Stoffel, M. (2019). The anomalous 2017 coastal el niño event in peru. *Climate Dynamics*, *52*(9), 5605–5622.

Rodríguez-Rodríguez, F., Pakomio Jara, O., Kuthe, N. M., Herrador-Colmenero, M., Ramírez-Vélez, R., & Chillón, P. (2019). Influence of distance, area, and cultural context in active commuting: Continental and insular children. *PloS one*, *14*(3), e0213159.

Salonen, M. (2014). *Analysing spatial accessibility patterns with travel time and distance measures: Novel approaches for rural and urban contexts* (Master thesis). University of Helsinski.

Sant'Anna, P. H., & Song, X. (2019). Specification tests for the propensity score. *Journal of Econometrics*, *210*(2), 379–404.

Saracostti, M., Lara, L., Martella, D., Miranda, H., Miranda-Zapata, E. D., & Reininger, T. (2019). Influence of family involvement and children's socioemotional development on the learning outcomes of chilean students. *Frontiers in Psychology*, *10*, 335.

Schlotter, M. (2011). *The effect of preschool attendance on secondary school track choice in germany-evidence from siblings* (tech. rep.). Ifo Working Paper.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). Dbscan revisited, revisited: Why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, *42*(3), 1–21.

Segretin, M. S., Hermida, M. J., Prats, L. M., Fracchia, C. S., Ruetti, E., & Lipina, S. J. (2016). Childhood poverty and cognitive development in latin america in the 21st century. *New directions for child and adolescent development*, *2016*(152), 9–29.

Smits, J., & Steendijk, R. (2015). The international wealth index (iwi). *Social indicators research*, *122*(1), 65–85.

Torche, F. (2014). Intergenerational mobility and inequality: The latin american case. *Annual Review of Sociology*, *40*, 619–642.

Trawick-Smith, J. (2019). Not all children grow up the same: Child development, diversity, and early care and education. In C. P. Brown, M. B. McMullen, & N. File (Eds.), *The wiley handbook of early childhood care and education* (pp. 29–58). John Wiley & Sons.

UNICEF (Ed.). (2019). *Children, food and nutrition* (The state of the world's children No. 2019). New York, NY, UNICEF.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health policy and planning*, *21*(6), 459–468.

White, H., & Sabarwal, S. (2014). Quasi-experimental design and methods. *Methodological briefs: impact evaluation*, *8*, 1–16.

Yachiyo Engineering. (2005). *Plan Maestro de Transporte Urbano para el Área Metropolitana de Lima y Callao en la República del Perú* (Final Report).

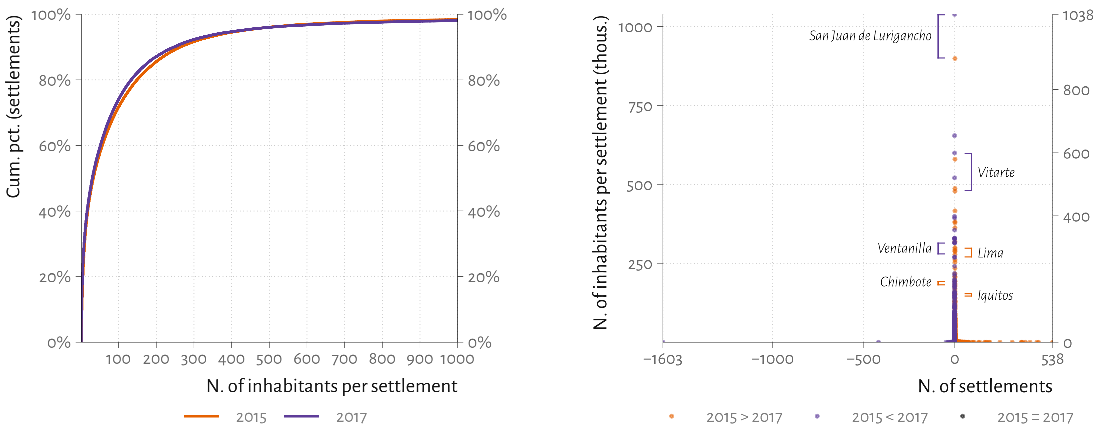
Yañez-Pagans, P., Martinez, D., Mitnik, O. A., Scholl, L., & Vazquez, A. (2019). Urban transport systems in latin america and the caribbean: Lessons and challenges. *Latin American Economic Review*, *28*(1), 1–25.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education. *Society for Research in Child Development*.

Yu, X., Ding, Y., Wan, W., & Thuillier, E. (2014). Explore hot spots of city based on dbscan algorithm, In *2014 international conference on audio, language and image processing*. IEEE.

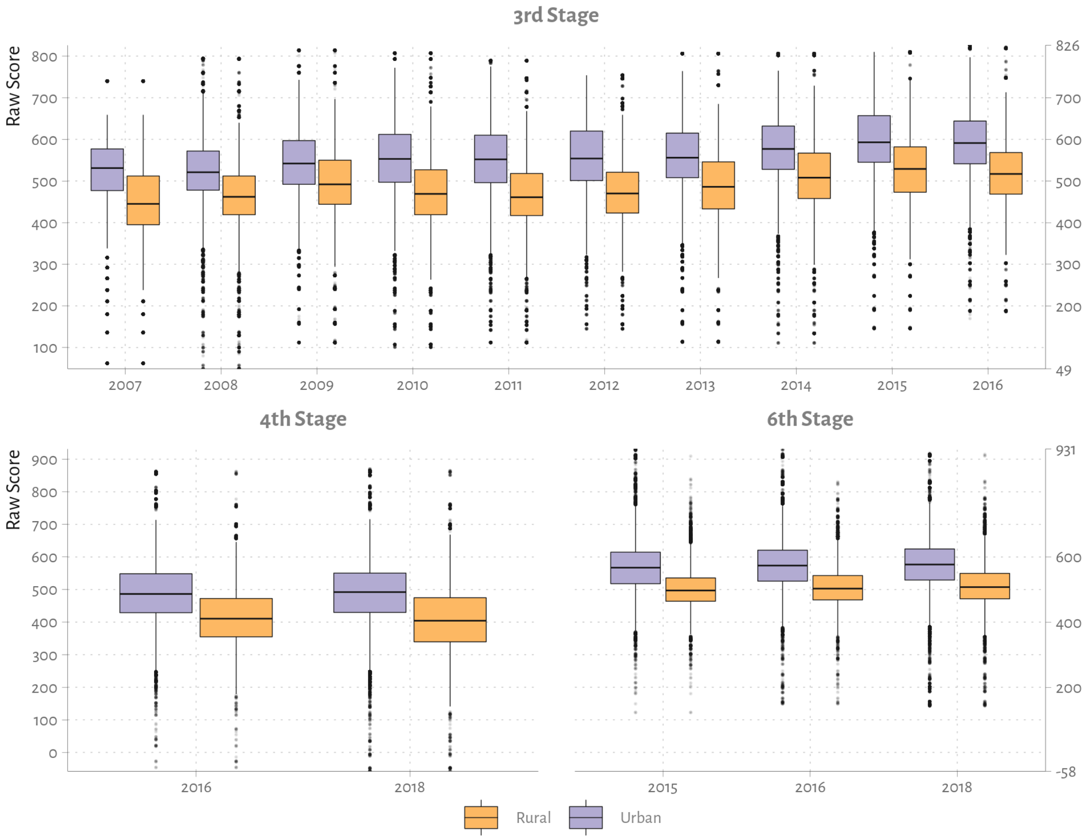
APPENDIX A
Supplementary Material

Figure A.1 Empirical Distribution of Population per Settlement.



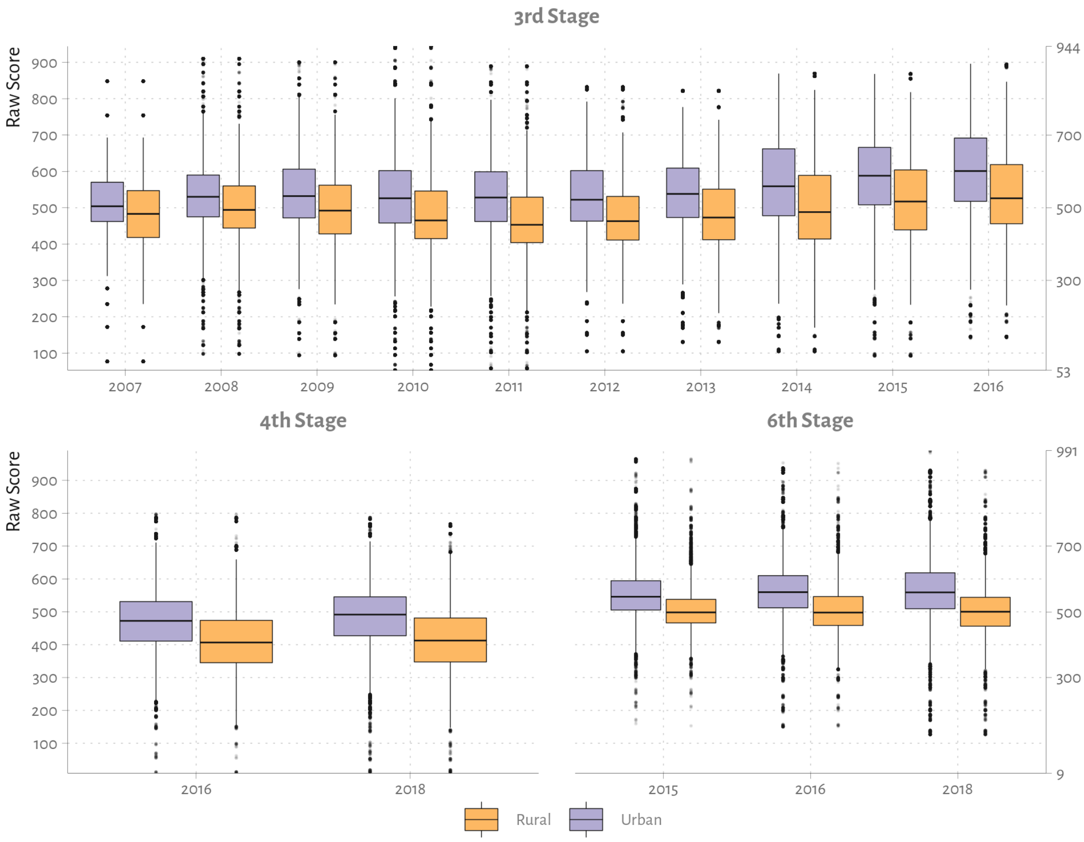
Source: Population and Housing Census, Peru 2017. *The right vertical axis is meant to help reading the graph, as it mirrors the left vertical axis.*.

Figure A.2 Language Test Scores by Living Area [ECE] (2007–2018).



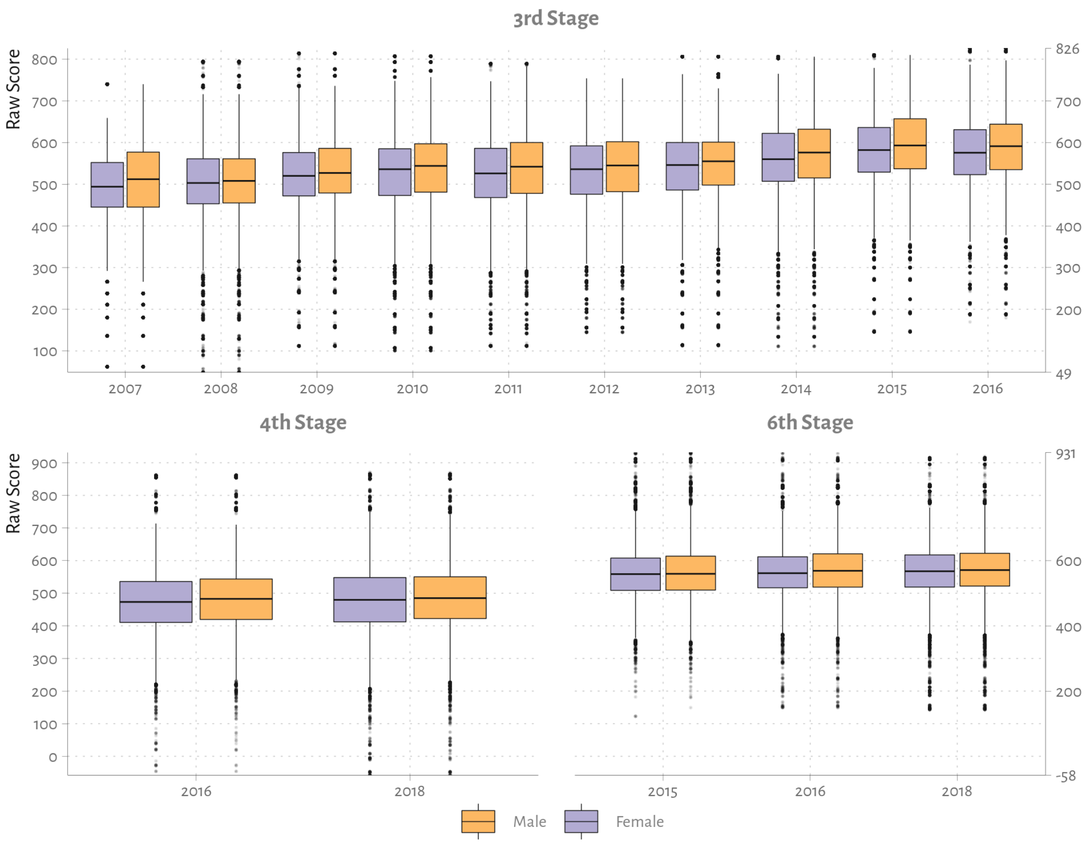
Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

Figure A.3 Mathematics Test Scores by Living Area [ECE] (2007–2018).



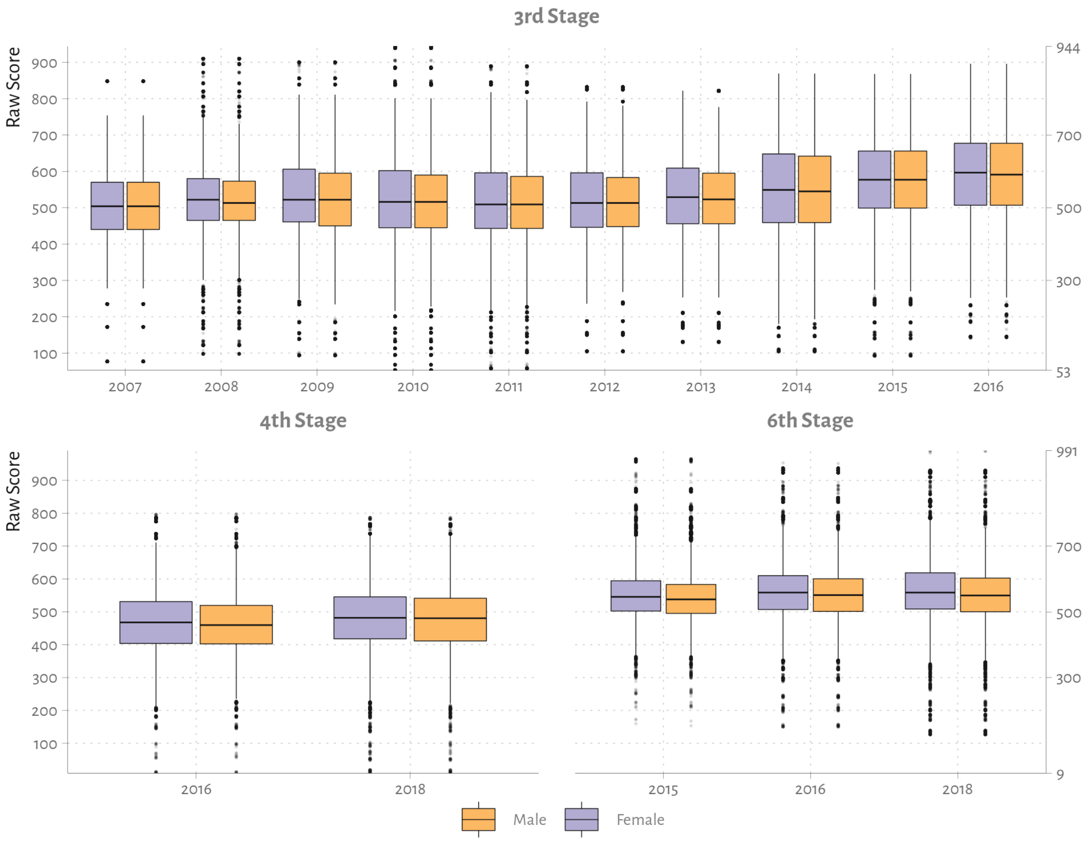
Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

Figure A.4 Language Test Scores by Sex [ECE] (2007–2018).



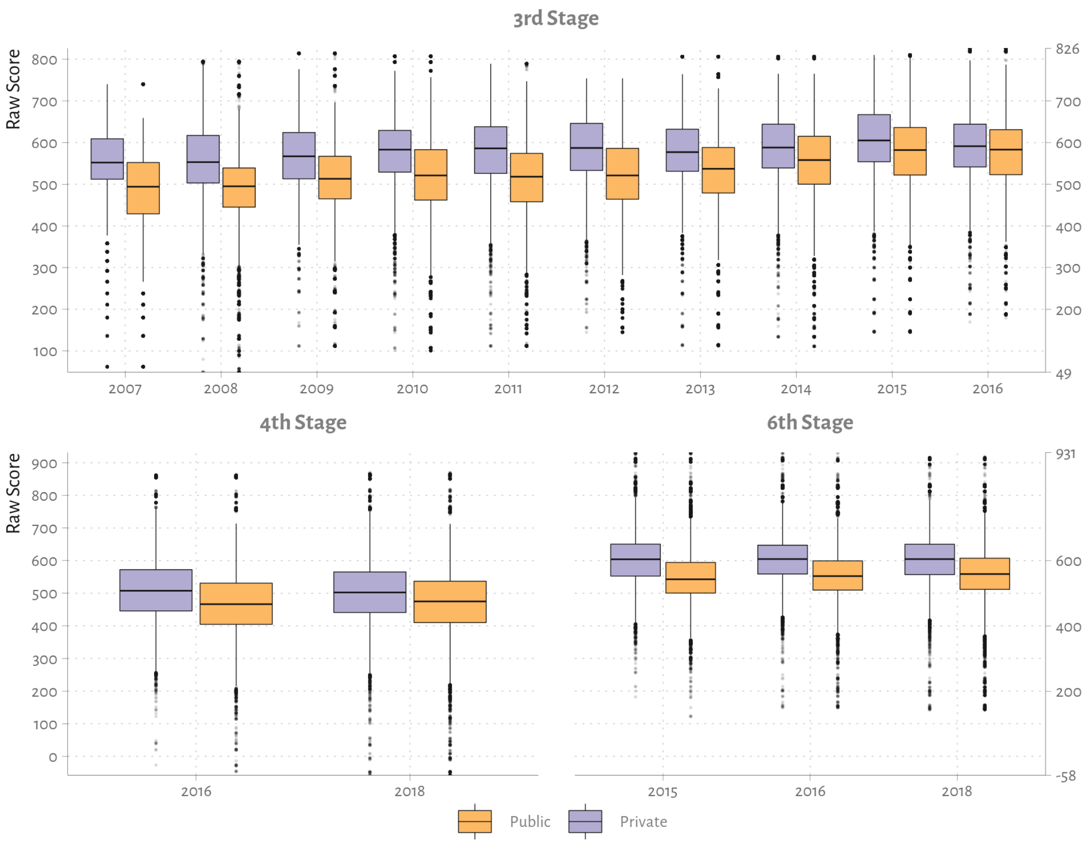
Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

Figure A.5 Mathematics Test Scores by Sex [ECE] (2007–2018).



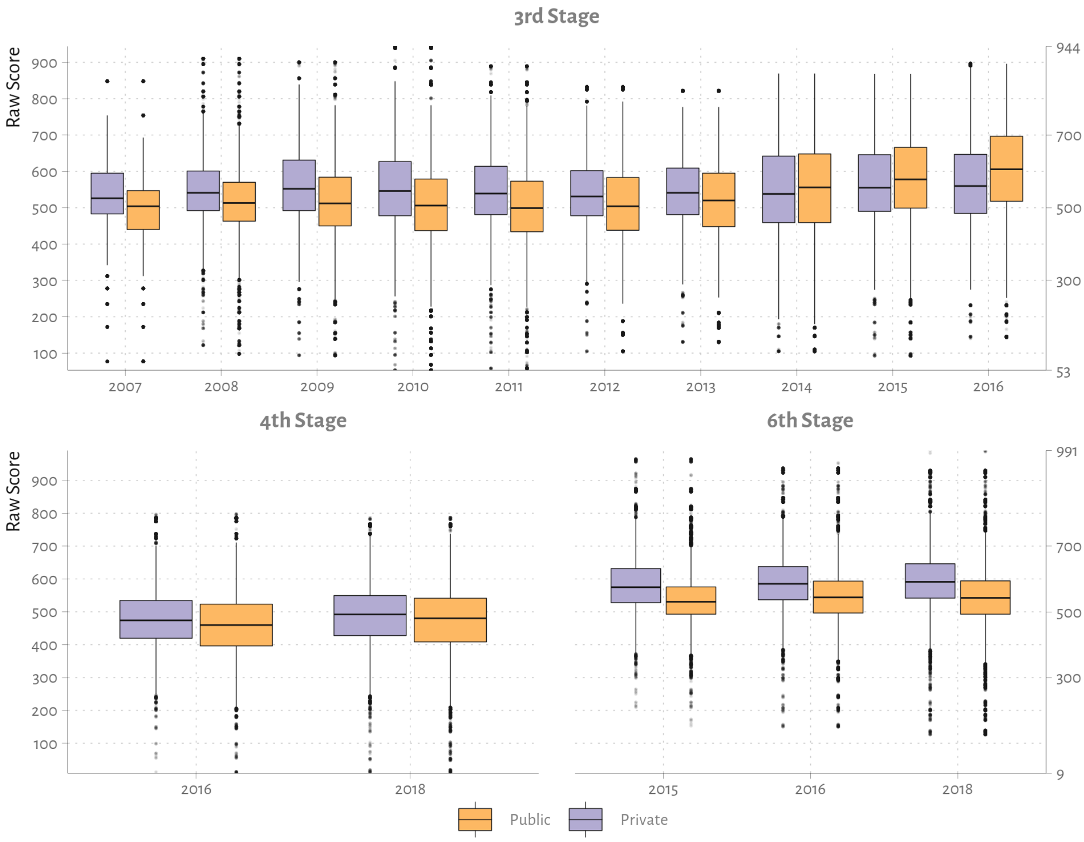
Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

Figure A.6 Language Test Scores by School Management [ECE] (2007–2018).



Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

Figure A.7 Mathematics Test Scores by School Management [ECE] (2007–2018).



Source: Oficina de Medición de la Calidad de los Aprendizajes: ECE. *The right vertical axis indicates the minimum and the maximum values of the shown series.*

I hereby confirm that the work presented has been performed and interpreted solely by myself except for where I explicitly identified the contrary. I assure that this work has not been presented in any other form for the fulfillment of any other degree or qualification. Ideas taken from other works in letter and in spirit are identified in every single case.

Bonn, October 15, 2021

Nicolas Eduardo Fajardo Acosta