UNIVERSIDAD DE CANTABRIA

PROGRAMA INTERUNIVERSITARIO DE DOCTORADO EN ECONOMÍA: INSTRUMENTOS DEL ANÁLISIS ECONÓMICO

TESIS DOCTORAL

Análisis económico de los efectos esperados y no esperados de políticas públicas en resultados educativos

Realizada por: Juan Guillermo Bedoya Ospina

Dirigida por: Marcos Fernández Gutiérrez

Escuela de Doctorado de la Universidad de Cantabria

Santander, Junio 2023

UNIVERSITY OF CANTABRIA

PhD Thesis

Economic analysis of the expected and unexpected effects of public policies in education outcomes

Author: Juan Guillermo Bedoya Ospina

Supervisor: Marcos Fernández Gutiérrez

Santander, June 2023

AGRADECIMIENTOS

El desarrollo de cualquier iniciativa investigativa que pretenda utilizar estándares de alta calidad en cada una de sus fases requiere contar con un contexto que así lo favorezca. Por lo tanto, esta tesis no hubiera sido posible sin la interacción con muchas personas e instituciones que permitieron que el contexto de desarrollo fuera lo más favorable posible.

En primer lugar, quiero agradecer al Dr. Marcos Fernández Gutiérrez, quién me acompañó en cada una de las fases de desarrollo de esta tesis, brindando espacios de discusión abierta y poniendo a mi disposición su tiempo, recursos y voluntad para lograr un trabajo riguroso. Desde nuestras primeras discusiones, logró identificar mis potencialidades y me brindó la confianza necesaria para comprometerme con llevar este proceso a buen puerto. El contexto también fue sumamente favorable desde mi familia, por eso agradezco a Viviana Carmona Lopera, Juanita Bedoya Carmona (quien nació en medio del desarrollo de este trabajo), a Luz Marleny Ospina Ruiz y a Ramón Octavio Bedoya. Gracias, familia, por permitir que nuestra vida familiar coincidiera con largas jornadas de lectura, exploración de datos, realización de documentos y viajes para presentar mis resultados.

La ciencia como proceso de construcción colectiva requiere encontrar individuos e instituciones que estén dispuestos a dedicar parte de su tiempo y sus recursos a aproximar las preguntas que se tenga de una manera metódica, a controvertir y escuchar a otros y, también, a aceptar los resultados derivados de las metodologías prevalentes en la comunidad científica. Esta tesis no hubiera sido posible sin el aporte y la generosidad de diferentes investigadores que vieron en mis ideas y trabajo un camino para la colaboración y construcción de conocimiento, agradezco entonces a Rafael de Hoyos, Ricardo Estrada, Fabio Sánchez, Mauricio Romero, Marcos Fernández-Gutiérrez, Gregorio Giménez, Mónica Yañez-Pagans, Marcela Silveyra, Elizabeth Jaramillo y Alejandra Quintana por su acompañamiento en este proceso. Así mismo, agradezco al Centro de Estudios de la Universidad de los Andes (CEDE), el Departamento para la Prosperidad Social (DPS), la Secretaría de Educación Pública de México (SEP), Panagora Group, USAID, Innovations for Poverty Action (IPA) y otras tantas instituciones que brindaron su apoyo de diferentes maneras para que esta tesis fuera posible.

Finalmente, quiero agradecer a la Universidad de Cantabria, y, en particular, a la Escuela de Doctorado, al Departamento de Economía y a sus profesores e investigadores por ofrecerme un espacio donde siempre encontré respuesta y apoyo a los diferentes retos que enfrenté en el camino de desarrollo de este trabajo.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Summary

The need to move towards a world where people who have lived historically marginalized are included and can be part of social prosperity is an imperative challenge for all those who are linked in one way or another to the world of economic development. In this context, education is presented as one of the main paths for social transformation, and, therefore, as one of the areas where governments, multilateral organizations, social organizations and academics, among others, have put their efforts and resources in recent years. Ensuring that children and young people attend school and obtain the necessary learning to be able to develop their potential is identified as a way to achieve greater effective and sustained social and economic inclusion. However, it is important to highlight that making investments does not ensure obtaining the expected results in educational terms. Some programs, projects and policies, even though they are designed and implemented by people committed to educational improvements and who use a large amount of economic resources, may not be having the expected impacts.

The objective of this thesis is to use impact evaluation techniques to understand from an economic perspective the expected and unexpected impact of programs, projects and policies on educational results. The confluence of elements such as the growth and availability of detailed datasets on living conditions and educational results, the emergence and consolidation of different impact evaluation techniques, and the existence of programs, projects, and policies that are being implemented with the aim of improving educational results or other relevant social indicators, create the propitious stage to explore what works and what does not through rigorous quantitative techniques.

This doctoral thesis is structured as follows: Chapter 1 presents the development context of this thesis, covering the motivation, methodological approach, research questions, and status in the publication process of the different chapters that compose it. Chapter 2 explores the impact on managerial skills

and student learning of a program implemented in Mexican public schools addressed at providing principals with skills for evidence-based decision-making and monitoring of activities and resources inside the schools they manage. The results reveal that the direct implementation through specialized professionals generated a positive impact on the managerial skills of the directors, however, no impacts are observed on student learning. Chapter 3 focuses on estimating the impacts on the cognitive abilities of new teachers derived from a reform of access to the civil service of public-school teachers in Mexico. The results reveal that the reform improved the cognitive abilities of the new teachers, an effect which was derived from an increase in the proportion of teachers who came from the meritocratic selection process, as well as a greater relationship between entrance tests and teachers' cognitive skills. As an unexpected effect, the reform did not manage to eliminate the discretionary entry route to the system and teachers entering through this route presented a very significant reduction in their cognitive skills. Chapter 4 focuses on estimating the impacts of the ceasefire within the framework of the Colombian peace process on the educational trajectory of the most vulnerable children and young people in the municipalities historically most affected by the Colombian armed conflict. The results obtained are unexpected, since, contrary to intuition, the decrease in armed actions increased the risk of dropping out for children and youth living in these municipalities. An exploration of the mechanisms reveals that this result was due to the increase in the cultivation of coca crops in the municipalities where armed actions were reduced, which ended up generating incentives for children and young people to drop out of school. Finally, Chapter 5 focuses on estimating the impact of the use of ICT at school on learning outcomes in mathematics and science. Using data from PISA 2018 for 35 countries, the results reveal that the association between use of ICT at school and learning outcomes is non-significant in most of the countries. This result is corroborated when considering only those countries for which a necessary condition for the causal interpretation of the results is met.

The different chapters contribute to the existing literature on the field of the economic evaluation of educational policies. Chapter 2 corresponds to the first experimental evaluation in a developing country of the impacts of a program aimed at improving the managerial capacities of school directors. Chapter 3 provides evidence to the literature on economics of education, labor economics and personnel economics on the importance of meritocratic selection in the profile of teachers. Chapter 4 is one of the first works on the effects of the de-escalation of armed activities on educational trajectories in contexts where other illegal activities are widely rooted in the territory. Finally, Chapter 5 is the first work that estimates the causal impact of the use of ICT in schools using the most well-known international comparable database on students' outcomes: PISA.

In terms of policy recommendations, the results of this thesis confirm that it is necessary to rigorously evaluate the efforts made by public policies to ensure that children and youth have better educational outcomes. The design and implementation of programs that use a large amount of resources does not ensure that the expected positive impacts are achieved and that unexpected negative impacts are avoided, so knowing what works and what does not will help to abandon some approaches and adjusting and/or scaling those with better results.

Resumen

La necesidad de avanzar hacia un mundo donde las personas que han vivido históricamente marginalizadas sean incluidas y puedan hacer parte de la prosperidad social es un reto imperativo para todos aquellos que, de una u otra manera, están vinculados al mundo del desarrollo. En este contexto, la educación se presenta como uno de los principales caminos para la transformación social, y, por ende, como una de las áreas donde gobiernos, organizaciones multilaterales, organizaciones sociales y académicos, entre otros, han puesto sus esfuerzos y recursos en los últimos años. Asegurar que los niños, niñas y jóvenes asistan a la escuela y obtengan los aprendizajes necesarios para poder desarrollar su potencial se identifica como un camino para lograr mayor una inclusión social efectiva y sostenida. Sin embargo, es importante resaltar que la realización de inversiones no asegura la obtención de los resultados esperados en términos educativos. Algunos programas, proyectos y políticas, aunque son diseñados e implementados por personas comprometidas con las mejoras educativas y que utilizan una gran cantidad de recursos económicos, podrían no estar teniendo los impactos esperados.

El objetivo de esta tesis es utilizar técnicas de evaluación de impacto para, desde una perspectiva económica, entender los impactos esperados y no esperados de programas, proyectos y políticas en resultados educativos. La confluencia de elementos como el crecimiento y disponibilidad de fuentes de información detalladas sobre condiciones de vida y resultados educativos, la aparición y consolidación de diferentes técnicas de evaluación de impacto y la existencia de programas, proyectos y políticas que se están implementando con el objetivo de mejorar los resultados educativos u otros indicadores sociales relevantes, crean el escenario propicio para explorar qué funciona y qué no a través de técnicas cuantitativas rigurosas.

La presente tesis doctoral se estructura como sigue. El capítulo 1 presenta el contexto de desarrollo de esta tesis, cubriendo la motivación, aproximación metodológica, preguntas de aprendizaje y estado en el proceso de publicación de los diferentes capítulos que la componen. El capítulo 2 explora el impacto en habilidades gerenciales y aprendizajes de los estudiantes de un programa implementado en escuelas públicas mexicanas con objeto de proveer a los directores con habilidades para la toma de decisiones, con base en evidencia y el monitoreo de las actividades y recursos al interior de las escuelas. Los resultados revelan que la implementación directa a través de profesionales especializados generó impacto positivo en las habilidades gerenciales de los directores; sin embargo, no se observan impactos en los aprendizajes de los estudiantes. El capítulo 3 se enfoca en estimar los impactos sobre las habilidades cognitivas de los nuevos profesores derivados de una reforma de acceso al servicio civil de los profesores de escuelas públicas en México. Los resultados revelan que la reforma mejoró las habilidades cognitivas de los nuevos profesores. Este resultado se derivó de un aumento de la proporción de los profesores que venían del proceso de selección meritocrático, así como de una mayor relación entre las pruebas de ingreso y las habilidades cognitivas de los profesores. Como un efecto no esperado, la reforma no logró eliminar la vía de ingreso discrecional al sistema y los profesores entrando por este camino presentaron una reducción muy significativa en sus habilidades cognitivas. El capítulo 4 tiene como objetivo estimar los impactos del cese al fuego, en el marco del proceso de paz colombiano, en la trayectoria educativa de niños, niñas y jóvenes más vulnerables de los municipios históricamente más afectados por el conflicto armado colombiano. Los resultados obtenidos son no esperados ya que, en contra de la intuición, la disminución de las acciones armadas aumentó el riesgo de deserción de los niños, niñas y adolescentes que habitan en estos municipios. Una exploración de los mecanismos revela que este resultado se debió al aumento de los cultivos de coca en los municipios donde las acciones armadas se redujeron, lo que terminó generando incentivos para que los niños, niñas y jóvenes dejaran la escuela. Finalmente, el capítulo

5 se enfoca a estimar el impacto del uso de nuevas tecnologías de la información y las comunicaciones (TIC) en la escuela sobre los aprendizajes de los estudiantes en matemáticas y ciencias. Utilizando datos de PISA 2018 para 35 países, los resultados revelan que la asociación entre uso de TIC y los resultados educativos no es significativa en la mayoría de los países. Este resultado se corrobora cuando se consideran solo aquellos países para los que una condición necesaria para la interpretación causal de los resultados se cumple.

Los diferentes capítulos realizan aportaciones relevantes a la literatura existente en el ámbito de la evaluación económica de políticas educativas. El capítulo 2 se corresponde con la primera evaluación experimental en un país en desarrollo de los impactos de programas orientados a la mejora de las capacidades gerenciales de los directores. El capítulo 3 brinda evidencia a la literatura de economía de la educación, economía laboral y economía del personal sobre la importancia de la selección meritocrática en el perfil de los servidores públicos. El capítulo 4 es uno de los primeros trabajos sobre los efectos de la desescalada de las actividades armadas sobre trayectorias educativas en contextos donde otras actividades ilegales están ampliamente arraigadas en el territorio. Finalmente, el capítulo 5 es el primer trabajo que aproxima el impacto del uso de las TIC en la escuela utilizando la base de datos comparable internacionalmente más relevante como es PISA 2018.

En términos de recomendaciones de políticas públicas, los resultados de esta tesis revalidan que es necesario evaluar de manera rigurosa los esfuerzos que se hacen para lograr que niños, niñas y adolescentes tengan mejores resultados educativos. El diseño e implementación de programas que utilizan gran cantidad de recursos no asegura que los impactos esperados se alcancen, de forma que conocer qué funciona y qué no, resulta fundamental para tratar de abandonar algunos enfoques y ajustar y/o escalar aquellos con mejores resultados.

Chapter 1. Introduction

1. 1 Motivation: Education as a vehicle for social transformation

The study and understanding of phenomena such as poverty, violence and social and political instability occupy an important part of the agenda in economics (Duflo and Banerjee, 2012; Acemoglu and Robinson, 2013; Ravallion, 2015). The last two centuries have shown clear advances worldwide in terms of some of the aforementioned phenomena, which results in the world population living in a time of unprecedented material prosperity (Deaton, 2013). However, it is also true that social progress has revealed new challenges, making urgent the understanding and effective attention to the material needs of historically excluded groups (Deaton, 2013). Recent decades have also seen the positioning of inequality as a central issue within social and political discussions, opening the debate on whether the path that led to unprecedented world prosperity is sustainable in a scenario of high economic inequality (Cingano, 2014).

In this context, which may have significant consequences on social and political stability in the coming years, different multilateral organizations, think-tanks, and other key institutional and social actors have highlighted the relevance of education as a vehicle for social transformation (United Nations, 2015; Narayan et al., 2018; OECD, 2023). The education sector is crucial on this regard, because significant improvements in education can lead to a reduction in inequality and discontent among some historically marginalized groups (Haveman and Smeeding, 2006). Governments have recognized the importance of achieving inclusive education systems, with high coverage and quality, as the main mechanism to combat poverty, low productivity and inequality (Cingano, 2014). In particular, in the case of Latin America, where inequality levels are the highest in the world (Hertz et al., 2007), the design and implementation of educational public policies is essential to achieve large-scale economic and social transformation.

Different countries around the world, in particular in Latin America, have opted for policies with a view to improving the access of the most vulnerable individuals to the educational system, as well as significantly increasing the quality of education these individuals receive (García and Saavedra, 2022). Conditional transfers, understood as cash transfers delivered periodically to the poorest families on the condition that the children belonging to these families attend school and frequent periodic medical check-ups, have played a very important role in the region (Stampini and Tornarolli, 2012). With a current poverty reduction scheme and improvement of human capital in the medium and long term, these types of programs have spread in the region, showing relevant positive impacts (García and Saavedra, 2022), but at the same time, drawing attention to the need for more specific policies to improve the socioeconomic conditions of the most vulnerable population (Millán et. al, 2019).

With the aim of promoting the quality of education received by the most vulnerable, different countries have designed and implemented programs and projects that focus on improving different factors associated with the quality of education (Viennet and Pont, 2017; Azevedo et al., 2021). These efforts included policies addressed at improving the profile of teachers who lead the educational processes in public schools where students with the most vulnerable profiles generally attend (FLACSO, 2019), improving the management of human and financial resources carried out by the principals of public schools (Bloom et. al, 2015), and promoting the use of information and communication technologies (ICT) in the educational process (Truncano, 2016), among other objectives, implemented in recent years in many countries across the world. Additionally, other types of programs or projects have been implemented that, although not focused on the education sector, have ended up having a direct or indirect influence on the educational results of the most vulnerable (chapter 4of this dissertation addresses a relevant case on this regard).

This dissertation is made up of a series of investigations that, incorporating a quantitative approach oriented to the generation of rigorous evidence in the field of the economics of education, focus on evaluating the expected and unexpected effects that different projects or public policies have had on educational results.

1.2 Methodological approach: Measurement of policy impact from an economic perspective as a way to find out what works and what does not

The last three decades have seen the convergence of different elements that are of interest for this dissertation. First, the importance of generating and compiling quality information on the living conditions of people, as well as on the actions carried out by governments, non-governmental organizations, and even private sector companies, has spread rapidly (OECD, 2018). The information available through surveys, administrative data and systematic multinational studies, among others, has grown significantly in recent years (OECD, 2018).

Second, different executors of policies and projects, academics, multilateral organizations and governments have increased their interest in having systematic and credible measurements that allow them to know the scope of their actions to improve the quality of life (ILO, 2018). The resources provided by donors and development financiers in the world are now often accompanied by the need for development and implementation of monitoring systems that make the resources and actions traceable within the countries or communities where they are used, in addition to evaluations with varied approaches that allow to communicate to the interested parties the results and/or impacts of the actions taken (Mitchell, 2021).

Third, methodological developments have appeared that are aimed at providing rigorous measurements regarding *what works and what doesn't*. In this last group, impact evaluation techniques have decisively broken into the world of economic fields such as development economics and economics of education, with the purpose of offering evidence on the impacts that different strategies, projects and policies, among others, have on dimensions such as the conditions of life of the most vulnerable (Gertler et al., 2016).

Impact evaluation techniques have permeated academia, multilateral development agencies, governments and even non-governmental organizations as a way to find out if the actions and resources allocated to development have attributable impacts, that is, if the beneficiaries of these resources and actions had observable changes in their living conditions attributable to the implemented actions (Cameron, Mishra and Brown, 2015; White and Raitzer, 2018). The objective of impact evaluation techniques then faces no lesser challenges in search of having *causal* measurements that allow isolating all those factors that surround the interventions in the world of development, as well as the people and communities in which these actions are carried out.

Impact evaluation techniques can be divided into two groups: experimental and quasi-experimental techniques (Gertler et al., 2016). The first group is oriented to bring the experimentalist approach to the social sciences. One of the main challenges of any impact measurement is to ensure that the measurement obtained is not the consequence of factors not directly attributable to what is intended to be measured, that is, that it is possible to distinguish between correlation and causality. The experimental techniques use the random assignment (through a lottery) of the actions to be evaluated as the way to ensure that the only difference between the groups that participate in an evaluation process is that some individuals, firms, administrative units, etc. receive an intervention by chance (i.e., a type of treatment) and others did not (Banerjee, Duflo and Kremer, 2016). In the simplest scenario, if randomization can be properly implemented, and, in addition, non-contamination can be maintained before the measurement periods, a simple mean difference between the treated and control groups can be interpreted as the causal impact of the treatment (for instance, a policy) is being evaluated.

The popularity of randomized controlled trials has grown in recent years, and even three representative authors for the world of development studies, Esther Duflo, Abhijit Banerjee and Michael Kremer, were awarded the Nobel Prize in Economics in 2019 for the "*experimental approach to alleviating global poverty*" (Bandiera, 2019). However, although the experimental method is currently recognized as the "gold standard" approach for impact evaluation, in some scenarios, this technique is not viable, for operational or political reasons, or simply because it does not fit the learning question a researcher wants to explore.

The second group of impact evaluation techniques are called quasi-experimental. Rather than using random assignment as the source of exogenous variation for the identification of causal effects, quasi-experimental techniques rely on the "natural" occurrence of exogenous variation (Gertler et al., 2016). The existence of observable eligibility conditions that generate cuts between different groups of individuals or the allocation of shares using clearly defined geographic areas, among other circumstances, generate scenarios where it is possible to identify causal impacts through quasi-experimental techniques. For example, the regression discontinuity approach uses the existence of cutoffs on unmanipulated selection variables as the origin of experiments in the neighborhood of this cutoff, while the follow-up of groups of individuals eligible to receive a particular program at different points in time or geographies allows the use of the difference-in-differences approach (Gertler et al., 2016).

1.3 Research questions: expected and unexpected effects of public policies on educational outcomes

This thesis explores the expected and unexpected impacts of various public policies on educational related results of different populations. This general objective is transferred to four themes, each corresponding to a chapter of the thesis and to relevant issues in the field of policies applied in recent years. The questions explored are the following:

- Do improvements in the managerial skills of public-school principals improve the learning of their students?
- Does a civil service entry reform change the cognitive skills of new teachers measured by standardized tests?
- What are the effects of the reduction of violence in areas with a high presence of illegal activities on the educational trajectory of students residing in those areas?
- What is the effect of the use of ICTs for learning at school on educational outcomes of students?

By answering these questions, this dissertation aims to contribute to the literature in the fields of economics of education and development economics in at least the following four dimensions. First, to provide evidence of the importance of using rigorous quantitative techniques to measure the impact of policies and projects in education. Second, to provide evidence that some policies and projects, in spite of using a large amount of resources and being led by officials and/or organizations with the interest of contributing to improve the living conditions of the most vulnerable, may not be having the expected positive impact and may have unexpected negative impacts. Third, to show the usefulness of administrative data and data collected at scale (i.e., Standardized tests) to answer questions related to economics fields such as economics of education and development economics.

And finally, to provide evidence that developing countries, and educational systems in general, continue to require the design, implementation and evaluation of innovative policies addressed at improving the quality of education that children receive.

1.4 Structure of this dissertation

This dissertation is made up of four essays that revolve around the expected and unexpected effects of policies and initiatives on educational outcomes. The first three of them (chapters 2 to 4) focus on Latin American countries (Mexico and Colombia) and make intensive use of administrative data to answer specific research questions posed on which are the impacts of efforts on school principals management skills, teachers' selection and peace achievement on educational outcomes. The last chapter (chapter 5) focuses on a sample of 35 countries from different continents and uses the results of a scaled standardized test to answer the research question posed on which are the expected and unexpected effects of policies oriented to increase ICT use at school on students' outcomes.

Chapter 2 uses a large-scale randomized experiment (in 1,198 public primary schools in Mexico) to study the impact of a policy which directly provided schools with high-quality management training from professional, in comparison to waterfall-style "train the trainer" model. The training (direct and waterfall-style) focused on improving principals' abilities to collect and use data to monitor students' basic numeracy and literacy skills and provide feedback to teachers on their instruction and pedagogical practices. After two years, direct training improved schools' managerial capacity relative to waterfall-style schools, but it had no impact on student test scores.

Chapter 3 studies the effect of a civil service reform that mandated the use of competitive test-based recruitment on the skill profile of new teachers in Mexico using a dynamic pre-post approach. Using administrative data from different sources and periods reaching a sample of about 25,000 new teachers, the results reveal that the reform led to the hiring of teachers with higher levels of cognitive skills and that this change was driven by an improvement at the bottom of the distribution of skills of the new hires. Two channels explain these effects. First, the reform reduced the prevalence of discretionary hires, which came disproportionately from the bottom of the skills distribution. Second, the reform improved the efficiency of hiring selection, making cognitive skills a more important determinant of hiring results. An unexpected effect of the reform was that the profile of teachers who entered through the discretionary route worsened in terms of non-cognitive skills.

Chapter 4 studies the effect of the ceasefire produced in the framework of the peace negotiations between the Colombian government and the FARC-EP guerrilla on the educational trajectories of the most vulnerable children and young people who live in the municipalities historically affected by the Colombian conflict. Using a difference-in-differences strategy, we found that the ceasefire increased the risk of dropping out school among the population of interest. We also found that the mechanism through which this effect occurs is the considerable increase in the cultivation of crops for illicit use, which created an incentive for student to abandon the education system. These results contribute to understanding the effects of the de-escalation of the conflict on human development trajectories when there are illicit economies that remain in the territory even after military actions have ceased.

Chapter 5 estimates the effect of both the use and the intensity of the use of information and communication technologies (ICT) for learning at school on student performance using internationally comparable data for about 170,000 students and 35 countries from PISA-2018. To do so, it takes advantage of within student variability in achievement and ICT use between two different subjects: math and science. The results show that the impact of both the use of ICT and the intensity of the use of ICT for learning at school on student performance in mathematics and science depends on the country. In most countries, however, the impact of both use of ICT and intensity of use of ICT on student performance is found to be not significant.

The four chapters of this thesis have been presented or are accepted to be presented at national and international seminars and conferences in the fields of economics of education and development economics. Additionally, they are in different states on the way to publication on top international journals in these fields (Table I.1).

Table I.1: Overview of the dissertation

Chapter title	**Events where it has been presented or is accepted**	**Publication status**
Direct vs indirect management training: Experimental evidence from schools in Mexico	1. Seminar by Foundation for Education and Development (FEDESARROLLO). Bogotá, Colombia. November 2019 2. Seminar School of Public Policy University of the Andes, Bogotá, October 2019	Published in *Journal of Development Economics*, 154, 1027-79
Rule–based civil service: evidence from a nationwide teacher reform in Mexico	1. Workshop at the University of the Basque Country. Bilbao, Spain. October 2022 2. Conference by Impact evaluation Network LACEA. Nashville, Tennessee, USA. April 2023 3. Seminar by Foundation for Education and Development (FEDESARROLLO). Bogotá, Colombia. May 2023 4. XXXI Meeting of the Economics of Education Association (AEDE). Santiago de Compostela, Spain. June 2023 5. Conference by International Economic Association (IEA), Medellín, December 2023 (accepted).	In the process of being published in the World Bank Policy Research Working Papers series.
Doing with One Hand and Undoing with the Other: Conflict De-escalation, Illicit Economies, and School Dropout Risk	1. XXVIII Meeting of Public Economics. A Coruña, Spain. May 2021 2. Seminar of the Department of Economics, University of Cantabria. Santander, Spain. June 2022	Under review in a Q1 JCR journal in Economics.

The impact of ICT use at school on student achievement: causal evidence from PISA based on variability within student	1. XXXI Meeting of the Economics of Education Association (AEDE). Santiago de Compostela, Spain. June 2023	In preparation for submission to a Q1 JCR journal

Chapter 2. Direct vs indirect management training in schools: experimental evidence from Mexico[ϕ]

2.1 Introduction

Schools are complex organizations that are often poorly managed. Across developed and developing countries, they tend to have worse management practices than hospitals and manufacturing firms (Bloom et al., 2014; Bloom et al., 2015). This is not surprising; school principals are chosen according to seniority in many countries. As a result, although they have years of classroom experience, principals may lack management skills.

We study the implementation of the Government of Mexico's large-scale *Escuela al Centro* (in English, school at the center) strategy, designed to strengthen school autonomy and improve principals' managerial capacity. This strategy was implemented nationwide for three consecutive school years (2015–16, 2016–17, and 2017–18). A core component was managerial training for principals that focused on collecting and using data to monitor students' basic numeracy and literacy skills and providing feedback to teachers on their instruction and pedagogical practices. We randomly assigned 1,198 eligible public primary schools to one of two groups: (1) a "train the trainer" group, which received managerial training under a cascade model in which 10% of school supervisors were trained by professional trainers, who then trained other supervisors, who in turn provided training to principals (n=599) and (2) a "direct training" group, in which principals received managerial training directly from a team of professional trainers (n = 599).[1]

[ϕ] This chapter is published as Romero, M., Bedoya, J., Yanez-Pagans, M., Silveyra, M., & de Hoyos, R. (2022). Direct vs indirect management training: Experimental evidence from schools in Mexico. *Journal of Development Economics*, 154. Some of the results included in this paper are part of Romero, M., Bedoya, J., Yanez-Pagans, M., Silveyra, M., & de Hoyos, R. (2021). School Management, Grants, and Test Scores: Experimental Evidence from Mexico. *Policy Research Working Paper*, *9535*. World Bank.

[1] Supervisors are the direct link between schools and educational authorities in each state. Supervisors are typically in

We collected data on schools' managerial practices, using the Development World Management Survey (DWMS) (Lemos & Scur, 2016), at baseline (in late 2015) and two years after the program was implemented (in early 2018). The DWMS measures different dimensions of schools' managerial practices, including operations management, people management, target setting, and monitoring. To measure students' learning, we use data from a nationwide standardized test (PLANEA).[2]

Our results show a significant improvement of 0.13 (p-value 0.018) standard deviations (σ thereafter) in managerial capacities among "direct training" schools compared to "train the trainer" schools. The improvements in managerial capacities do not translate into meaningful impacts on student learning. Students in "direct training" schools have test scores that are 0.03σ higher than their counterparts in "train the trainer" schools. However, this difference is not statistically significant (p-value 0.24) and we can rule out an effect greater than 0.08σ on test scores at the 95% level. There is little evidence of heterogeneity in treatment effects by baseline school characteristics.

The failure of "direct training" to significantly improve learning outcomes could be related to the weak contemporary correlation between managerial practices and test scores in Mexico (as measured by the DWMS). Our baseline data shows that a 1σ improvement in managerial practices is associated with an increase of less than 0.1σ in test scores, a weaker correlation than Bloom et al. (2015) reported for several countries. However, even assuming a stronger link between management and test scores (an increase of 0.4σ in test scores as the management index increases by one standard deviation) based on the results from Bloom et al. (2015) would imply that an increase of 0.13σ in management practices should yield an increase in test scores of 0.029σ—the actual treatment effect was 0.03σ.[3]

---

charge of 8 to 20 schools (Santiago et al., 2012).

[2] *Plan Nacional para la Evaluación de los Aprendizajes* (PLANEA) was designed by the Mexican Education Evaluation Institute, which measures Math and Spanish learning outcomes in grades 6, 9, and 12. PLANEA is aligned with the national curriculum and applied to a sample of students in all Mexican schools. In schools with fewer than 40 students in the grade assessed, every student is tested. In those with more than 40 students, a random sample is tested.

[3] Alternately, using our own data—and under some strong assumptions that allow us to use the treatment as an instrument for DWMS scores—our treatment effect on DWMS implies an expected increase of .065σ in test scores, given the

Overall, the expected treatment effects on learning outcomes (assuming previous correlational evidence is causal and given the treatment effects on management practices) are of the same order of magnitude as the actual treatment effects. While the intervention improved management practices, these improvements did not generate statistically significant (even with a sample size of 1,198 schools) changes in learning outcomes. The fact we do not find treatment effects on test scores is not due to a lack of power. Our ex-post minimum detectable effect (MDE) is 0.081σ for test scores (with power of 80% and size of 5%) (Ioannidis, Stanley, & Doucouliagos, 2017; McKenzie & Ozier, 2019). Rather, this result likely implies the need for larger effects on management practices to find economically meaningful effects on test scores.[4]

One way to boost the intervention's impact on management practices would be to increase principals' attendance to the training workshops. While "direct training" principals were about ten percentage points more likely to complete courses or receive counseling on how to carry out school director duties in the past, less than 25% completed the entire training (~ 80 hours), and roughly 10% completed less than 20 hours of the training. Instrumental variable approaches suggest boosting attendance to the training workshops would result in further improvements in management that would translate into meaningful impacts in student learning outcomes. However, we take these results as suggestive evidence that requires further confirmation in future studies due to measurement errors in the attendance data.

We contribute to the literature and policy debate on improving school management in low- and middle-income countries. Our study advances research that explores the relationship between school management and student outcomes (World Bank, 2007). Recent evidence, mostly from developed

treatment effect on DWMS scores.

[4] Alternatively, it could be the case that schools in Mexico are so well managed that the returns to additional increases in management are relatively low. However, comparing the distribution of DWMS scores in our setting to those in other countries found by Bloom et al. (2015) suggests this is not the case.

countries, demonstrates that management practices are an important determinant of school effectiveness. Using data for 39 charter schools in the United States, Dobbie and Fryer (2013) show that traditional school inputs such as class size and teaching certifications cannot explain differences in school effectiveness. However, school management practices, such as providing feedback to teachers and using data to guide instruction, are a significant determinant of school effectiveness (Fryer, 2014). In line with Fryer (2014)'s findings, Bloom et al. (2015) document a positive and statistically significant correlation between managerial practices and student learning outcomes. There is also evidence from India that learning outcomes and progress are positively correlated with managerial practices (Lemos, Muralidharan, & Scur, 2021). Our baseline data adds to the evidence base on the correlation between school management and learning outcomes. We find a weaker correlation between them than previous studies have identified, which could be partially explained by the low autonomy in the Mexican public education system — Bloom et al. (2015) shows higher school autonomy is correlated with higher management scores.[5]

Moreover, we provide experimental estimates of the relative effectiveness of two strategies to improve school principals' managerial capacity on management practices and student learning outcomes in a developing country. While there is evidence from the US that training programs to improve school principals' managerial practices have a positive effect on student learning outcomes (Fryer, 2017), our evidence and findings from other developing countries suggest otherwise. A closely related paper by Muralidharan and Singh (2020) shows that an attempt to improve management quality in Indian schools by inducing principals to adopt "best practices" had no impact on student outcomes. India's accountability and incentive structure for principals is rather weak (as it is in Mexico), which the authors argue may explain why improving managerial practices has little

[5] Mexican schools are less autonomous than schools in other Organization for Economic Co-operation and Development (OECD) countries (Hopkins et al., 2007; OECD, 2016).

or no effect on test scores.[6]

2.2 Context and intervention

Context

Mexico's primary education system (grades 1 to 6) has more than 14 million students and 573,000 teachers distributed across roughly 100,000 schools.[7] The system is highly decentralized: 32 state-level education systems follow a common national curriculum and general guidelines from the Federal Secretariat of Public Education (Federal SEP, from its acronym in Spanish). However, local governments are fully responsible for administering each state-level Secretariat of Public Education.

Access to primary education in Mexico is high, with over 98% of children aged 6 to 12 enrolled in the education system (World Bank, 2017b; Dirección General de Planeación, Programación y Estadística Educativa, 2018). However, the quality of education is low. Although almost all children graduate from primary school (World Bank, 2017a), fewer than half of them achieve basic proficiency in math and Spanish (and only one in three in marginalized areas) according to 2018 nationwide standardized tests (Instituto Nacional para la Evaluación de la Educación, 2018).

Mexico has three types of public primary schools: general primary schools (which teach most children), and indigenous and community schools, which serve roughly 800,000 and 400,000 students, respectively. These include many small, multi-grade schools with small numbers of students.[8] The existence of a large number of small schools increases the governance challenges and requires tailored management models. These governance challenges are compounded by a high

---

[6] A second potential explanation for the lack of impact is that managerial practices take longer to improve student education outcomes (see de Hoyos, Ganimian, and Holland (2020)).

[7] Unlike other countries in Latin America, Mexico has a small private education sector that accounts for only 10% of the total primary enrollment (Elacqua, Iribarren, & Santos, 2018).

[8] The smallest 40% of primary schools in the country serve 8.5% of its primary school students. By comparison, Mexico has less than half of the student population of the United States, but 50% more schools.

rotation of teachers and school principals and—until recently—the lack of a system to regulate the entry and promotion of teachers. Previously, the national teachers' union influenced teachers' (and school principals') appointments (Álvarez, García-Moreno, & Patrinos, 2007). In 2013, the central government implemented a major education reform that defined and regulated a merit-based process to hire and promote teachers and principals. It also introduced the Escuela al Centro strategy to enhance principals' managerial capacities to improve students' learning outcomes.

The Escuela al Centro strategy

The government implemented the Escuela al Centro strategy nationwide for three consecutive school years (2015–16, 2016–17, and 2017–18). It had two main components: the provision of school grants and school principals' managerial training.[9] The grant component consisted of an annual monetary transfer to schools that submitted an improvement plan approved by their school council. The grants ranged from USD 1,500–15,000 depending on the school's size (about USD 5–50 per student). Schools used these grants to implement their annual improvement plans and pay for basic supplies and repairs. All schools in our sample received these grants.

The training component focused on improving school principals' capacity to collect and use data to monitor students' basic numeracy and literacy skills and provide teachers with feedback on their teaching styles. To implement this training, the Federal SEP developed two tools: (i) a student assessment to monitor foundational skills (Sistema de Alerta Temprana en Escuelas de Educación Básica, SisAT) and (ii) a Stallings classroom observation tool to provide feedback to teachers on how to improve their instructional and pedagogical practices.

The SisAT was developed based on evidence that providing school principals in Mexico with

[9] The description of the Escuela al Centro strategy is available at: http://www.dof.gob.mx/nota detalle popup.php?codigo=5488338,
and the operating rules are available at: http://www.dof.gob.mx/ nota detalle.php?codigo=5509544&fecha=29/12/2017.

information on what areas of the national curriculum are the most challenging for students, based on national standardized learning assessments, had positive effects on student learning (de Hoyos, García-Moreno, & Patrinos, 2017; de Hoyos, Ganimian, & Holland, 2019). It includes items from past national standardized assessments to measure students' basic numeracy and literacy skills and identify lagging students to trigger early remedial actions. Teachers administer the SisAT and input the scores into a simple software program that generates a detailed report and flags students with significant learning gaps. The SisAT also pinpoints the most challenging areas of the national curriculum for students and classrooms. While schools were free to decide when to administer the SisAT, most did so at the beginning of the school year to generate baseline measures to include in their school improvement plans and throughout the school year to monitor students' progress.

The Stallings classroom observation tool was developed based on evidence that using school principals to coach teachers improves student learning in Mexico (Secretaría de Educación Pública & Banco Internacional de Reconstrucción y Fomento, 2015). It collects information on the teacher's use of time in the classroom, including the activities conducted, pedagogical practices, use of educational materials, and level of student engagement (Stallings, 1977; Stallings & Molhlman, 1988). The tool helps school principals systematically collect data to provide feedback to teachers on how to improve their teaching.

The Federal SEP developed a high-quality training strategy, including learning materials, to help principals use the SisAT and the Stallings classroom observation tool. The training consisted of 40 hours of instruction per tool.[10] The SEP used a "train the trainer" cascade model to roll out the Escuela al Centro strategy throughout the country. State-level education authorities selected 10% of all primary school supervisors to receive the training from a professional team that included staff

[10] These training materials are available at the Escuela al Centro website: https://escuelaalcentro.com/intervenciones/descarga-los-materiales/.

involved in designing the tools. The trained supervisors then provided training to the other supervisors in their state. After all supervisors in a state were trained (by either the professionals or their peers), they then proceeded to train the school principals in their jurisdictions. To test the efficacy of the cascade model versus professional training, the SEP provided professional training to some school principals.

2.3 Research design and data

Sampling and randomization

To test the effectiveness of the professional training, the SEP invited all 32 states in the country to participate in an impact evaluation. The seven states that met the requirements— Durango, Estado de México, Morelos, Tlaxcala, Guanajuato, Tabasco, and Puebla—were selected to be part of this research study (see Figure A.1).[11]

The local education authorities invited all public primary schools in all seven states to apply for the school grant component of Escuela al Centro. We randomly assigned the 1,198 schools that applied to the grants to one of two groups: (1) "train the trainer" schools, which received a school grant and school principals' managerial training using the cascade model (n = 599) or (2) "direct training" schools, which received a school grant and professional training (n = 599).[12]

Our experimental design allows us to estimate the causal effects of using professional trainers vs. the cascade model to train school principals.[13] Our sample included public primary schools that chose to

---

[11] From the 32 states in Mexico, 14 states expressed interest in participating in the impact evaluation. However, only seven complied with the required paperwork.

[12] Some principals in "direct training" schools also benefited from short-term leadership certificate training programs offered by state-level education authorities. These programs focused on leadership issues, in line with the national school principal's profile standards.

[13] While it is not possible to experimentally identify the impact of the cascade-style training vis-a`-vis no training at all, there is evidence that cascade training models tend to be relatively ineffective (Popova et al., 2018).

participate in the program. To be eligible, schools had to have more than 60 students; those with at least one classroom with students from different grades were excluded.[14] Therefore, the schools included in the experiment have more students and teachers and are more likely to be urban than the average public primary school in Mexico (see Table A.1).

The randomization protocol varied slightly across the seven participating states. Broadly, schools were first stratified into different groups (by enrollment and location) and then randomly assigned to either the treatment ("direct training" by professional trainers) or control (cascade-style training) group.

Data

We collected primary data on the principals' managerial practices and perceptions of the quality of the training they received. We also use secondary data from administrative records provided by SEP that include: (i) student learning outcomes; (ii) school marginalization index; and (iii) information on schools' infrastructure, enrollment rates, and number of teachers. Our study period coincides with two school years, 2015–16 (baseline) and 2017–18 (follow-up). In addition, the baseline and follow-up months roughly coincide with the nationwide standardized test application dates, which allow us to measure the intervention's impact on both management practices and student test scores.

Primary data

Information on schools' managerial practices was collected using the DWMS—an adaptation of the World Management Survey (WMS), originally developed to measure the quality of management practices in manufacturing firms in developed (Bloom & Van Reenen, 2007) and developing countries (Bloom et al., 2013).[15] The WMS and the DWMS were subsequently adapted to measure

[14] Small schools were excluded because the managerial intervention was focused on training principals to coach teachers. In small schools, principals also teach and thus need different management models.

[15] For more on the DWMS survey instrument, see Lemos and Scur (2016) and https:// developingmanagement.org/

management quality in the education and health sectors (Bloom et al., 2015; Bloom at al., 2015a). The WMS and DWMS are fully comparable; the latter can better identify granular differences in management practices at the lower end of the management quality distribution, where most public schools and hospitals in developing countries are located.

The DWMS adaptation to measure management practices in schools in developing countries consists of a recorded interview with the school principal. The interview includes 23 open-ended questions that collect information on four dimensions: operations management, people management, target setting, and monitoring.[16] The interviews, conducted by a team of two trained enumerators (one coder and one interviewer), lasted around two hours. While the DWMS is designed to be less subjective than the WMS to overcome the lower capacity of enumerators in developing countries, there is still considerable room for enumerator subjectivity in data coding. We assigned the same team of trained enumerators to code the audio files from all the original interviews to ensure comparability over time. Unfortunately, 32% of the audio files from the baseline, and 16% from the follow-up, were damaged when we asked the enumerators to code the interviews. Schools with and without misplaced audio files in the endline are statistically indistinguishable in observable characteristics (see Tables A.2 and A.3). Thus, our results are unlikely to be driven by differences in observable or unobservable characteristics between schools with and without functioning audio files, including the treatment status. To ensure comparability across schools, we randomly assigned audio files to enumerators and control for enumerator fixed effects in all the regressions. We conducted the baseline DWMS surveys between October 2015 and May 2016, and the follow-up surveys from January to May 2018.[17]

[16] The DWMS adaptation for Mexico included an additional dimension, leadership. Having this additional dimension responded to the government's need to better align the DWMS instrument to the rules of operation of Escuela al Centro. All the analyses reported in this paper exclude the leadership dimension when constructing the overall DMWS index to ensure it is comparable with other settings.

[17] https://escuelaalcentro.com/ has a detailed timeline of when different rounds of data collection took place in each state.

For reference, we compare the distribution of management scores in our setting (at baseline) to the distribution in India, Brazil, and the US from Bloom et al. (2015)— see Figure A.2. Overall, the average school in our setting has a higher management score than the average school in India (2.1 vs 1.7), a similar score to the average school in Brazil (2.1 vs 2.0) and a lower score than the average school in the US (2.1 vs 2.7). However, the dispersion in management practices in our setting is lower, which could be explained by the restrictions imposed on the experimental sample (e.g., excluding small multi-grade public schools and all private schools).

School principals also completed two online surveys to assess the quality of managerial training—one for each tool. The surveys included questions about different elements of the tools and their associated training. Since the surveys were not mandatory, many school principals did not complete them. Schools that answered the online surveys are statistically different from those that did not in several observable characteristics, including the treatment status (see Tables A.9–A.12). For completeness, we report some basic statistics from these two online surveys. However, their information is not representative of our experimental sample due to sample selection (i.e., it has differential attrition across treatments and within each treatment); therefore, we exclude this data from our main analysis.

Secondary data

We use three types of secondary data. First, we measure student learning outcomes using PLANEA test scores. The exam was administered to grade 6 students in June 2015 and June 2018. SEP gave the authors access to anonymized student-level data for both years for all schools in our sample. As part of registering their school for PLANEA, principals need to fill a survey (PLANEA-Contexto). The survey asks about their daily activities and the challenges they face. We use these surveys as a secondary measure of principals' management practices and their exposure to the training.

Second, we gathered information on the location of each school from the PLANEA data. We used this information to match each school to its locality's marginalization index, which accounts for deficiencies in education, housing, population, and household income.[18] Third, we use administrative school census data collected by federal and state-level education authorities known as *Formato 911*. Since 1998, *Formato 911* has been collected at the beginning and end of each school year. It gathers basic information on the number of students, the number of teachers and their qualifications, the school principal's characteristics, the number of classrooms, and its geographic location. This school census data can be matched with the PLANEA data.[19]

Balance and attrition

Most student and school characteristics are balanced across treatment arms at baseline (see Table 2.1). The average school in our sample has 279 students, 9.4 teachers, and a pupil–teacher ratio of 29; 40% of schools are in rural areas and 38% are in areas categorized as poor or very poor by the government. The last two rows of the table show the fraction of schools for which we have endline DWMS and PLANEA data (in 2018). We have PLANEA data for nearly all schools (~99%) and DWMS data for ~77% of schools (due to damaged audio files from the interviews, as mentioned above). The proportion of schools with both PLANEA and DWMS data is balanced across treatments.

[18] Consejo Nacional de Población (CONAPO) estimates this index.

[19] All the data used in this paper can be downloaded from www.xaber.org.mx.

Table 2.1: Balance across treatment groups

	(1)	(2)	(3)
	Mean (SD)		Difference
	Train the trainer	Direct training	(2)-(1)
Students in math achievement L-IV (%)	7.79	8.36	0.56
	(11.11)	(12.25)	(0.66)
Students in math achievement L-I (%)	60.00	60.17	0.19
	(21.81)	(22.24)	(1.22)
Students in language achievement L-IV (%)	2.67	3.31	0.65**
	(3.86)	(6.40)	(0.30)
Students in language achievement L-I (%)	52.17	51.56	-0.60
	(20.25)	(20.52)	(1.15)
Marginalization	0.38	0.38	-0.00
	(0.49)	(0.49)	(0.02)
Urbanization	0.41	0.39	-0.02
	(0.49)	(0.49)	(0.02)
Number of students	272.59	285.96	13.31
	(163.74)	(163.69)	(8.87)
Number of teachers	9.27	9.63	0.36
	(4.23)	(4.39)	(0.24)
Student-teacher ratio	28.34	28.89	0.54
	(6.92)	(7.18)	(0.35)
DWMS endline missing	0.22	0.23	0.01
	(0.41)	(0.42)	(0.02)
PLANEA endline missing	0.01	0.01	0.00
	(0.08)	(0.09)	(0.00)
Observations	599	599	1,197

This table presents the means and standard deviations (in parentheses) for "train the trainer" (Column 1) and "direct training" schools (Column 2). The differences reported in Column 3 take into account the randomization design (i.e., including strata fixed effects), and standard errors (in parentheses) are clustered at the school level. Achievement level (L) refers to the PLANEA 2015 exam results, which are scored from L-I (lowest) to L-IV (highest). *Marginalization* is a variable coded 1 for areas with "high" or "very high" marginalization, and 0 otherwise according to CONAPO. *Urbanization* is a variable coded 1 for schools located in an urban area, and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the 2015–2016 academic year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Compliance

To measure compliance with the evaluation's original design, we compiled information on whether school principals reported attending the training sessions on the two tools. As mentioned above, since the characteristics of schools that answered the survey are different from those that did not (see Tables A.9–A.12), these results should be interpreted with caution. Due to the sample selection in the compliance measures and the inability to directly compare the training hours across treatment arms (cascade vs. direct), local average treatment effect estimates using the treatment assignment as an instrument for the number of training hours principals report are difficult to interpret and likely biased. While virtually no principals in "train the trainer" schools completed the full training on the use of either tool, less than half (~40%) received some training (10–39 hours) through the cascade model (see Columns 1 and 2 of Table 2.2). About one-quarter of principals in "direct training" schools (20–25%) completed the training on both tools, and roughly 80% received some training from professionals. The difference between treatment groups is statistically significant for both completed training and the indicator for some training. This is further supported by evidence from surveys principals completed as part of the nationwide student standardized test (PLANEA-Contexto surveys) in 2018. Specifically, "direct training" principals were more likely to complete courses or receive counseling on how to carry out school director duties in the past 12 months (see Panel A, Table A.4)

Table 2.2: Compliance across treatment groups

	(1)	(2)	(3)
	Mean (SD)		**Difference**
	Train the trainer	Direct training	(2)-(1)
Panel A: Stallings classroom observation tool			
All training sessions (40 hours)	0.01	0.24	0.23***
	(0.10)	(0.43)	(0.02)
Some training sessions (10-40 hours)	0.39	0.86	0.44***
	(0.49)	(0.35)	(0.03)
Observations	304	533	837
Panel B: Foundational skills measurement tool (SisAT)			
All training sessions (40 hours)	0.01	0.19	0.18***
	(0.09)	(0.39)	(0.02)
Some training sessions (10-40 hours)	0.32	0.72	0.39***
	(0.47)	(0.45)	(0.03)
Observations	402	464	866

This table presents the means and standard deviations (in parentheses) for "train the trainer" (Column 1) and "direct training" schools (Column 2). The differences reported in Column 3 take into account the randomization design (i.e., including strata fixed effects), and standard errors (in parentheses) are clustered at the school level. Panel A has information on whether the school principal attended the training sessions for the Stallings classroom observation tool (and how many hours). Panel B has information on whether the school principal attended the training sessions on SisAT (and how many hours). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.4 Results

Correlation between management (DWMS) and learning

We first explore the correlation between learning outcomes and DWMS at baseline. We seek to replicate the analysis in Bloom et al. (2015) and compare our results with those previously found in the literature on the magnitude of the relationship between student learning outcomes and school management measured by the DWMS.

In our data, better management quality, as measured by the DWMS, is only marginally correlated with better educational outcomes (see Table 2.3). A one-standard-deviation increase in the DWMS index is associated with an increase of 0.00σ–0.02σ in student test scores. We follow Bloom et al.

(2015) and control for the number of pupils in the school, the pupil-teacher ratio, and the marginalization index (Column 4). We also control for measurement error by adding interviewer fixed effects (Column 5). The point estimate is robust to various controls and is never statistically significant. By comparison, Bloom et al. (2015) find that a one-standard-deviation increase in the WMS index is associated with an increase in pupil outcomes of 0.2–0.4σ. In Brazil, the setting included in their study closest to Mexico, a one-standard-deviation increase in the WMS index is associated with an increase in pupil outcomes of 0.104σ. Thus overall, we find a lower correlation between outcomes and management than previously documented in other countries.

Of the four components of the DWMS (operations, monitoring, targets, and people), "Targets" was the most closely correlated with student outcomes, followed by "Monitoring and people"; none of them demonstrated a statistically significant correlation with test scores in our setting (see Table A.5).

Table 2.3: Association between DWMS and test scores at baseline (all schools in the sample)

	(1)	(2)	(3)	(4)	(5)
	PLANEA 2015 scores				
DWMS	0.0017	0.011	0.020	0.017	-0.0065
	(0.025)	(0.025)	(0.023)	(0.022)	(0.027)
No. of obs.	20,680	20,680	20,680	20,049	20,049
State FE	No	Yes	Yes	Yes	Yes
Strata FE	No	No	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes
Enumerator FE	No	No	No	No	Yes

This table presents the conditional correlation between the DWMS and student test scores at baseline across all schools in our sample. State FE indicates whether state fixed effects are included. Strata FE indicates whether strata fixed effects are included. Controls indicates whether the regression controls for the number of pupils in the school, the pupil– teacher ratio, and the marginalization index. Enumerator FE indicates whether interviewer dummies are included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Experimental results

Our main estimating equation for student-level outcomes is:

$$Y_{isg} \quad = \quad \alpha_g + \gamma_1 DirectTraining_s + \varepsilon_{isg} \qquad [2.1]$$

where Y_{isg} is the outcome of interest of student i in school s in group g (denoting the stratification group used to assign treatment), α_g are strata fixed effects, $DirectTraining_s$ indicates whether school s received training directly provided by professional trainers, and ε_{isg} is an error term. We use a similar specification without i subscript to examine school-level outcomes. We estimate these models using ordinary least squares, clustering standard errors at the school level. γ_1 is the coefficient of interest and reflects the difference between the two types of training.

Table 2.4: Effects on the DWMS and on learning outcomes

Panel A: DWMS and its components

	(1) DWMS	(2) Operations	(3) Monitoring	(4) Targets	(5) People	(6) Leadership
Direct training	0.13^{**}	0.14^{**}	0.13^{**}	0.027	0.093^{*}	-0.0091
	(0.053)	(0.056)	(0.060)	(0.052)	(0.056)	(0.060)
No. of obs.	913	913	913	913	913	911

Panel B: Learning outcomes

	(1) Math	(2) Language	(3) Average	(4) PCA
Direct training	0.031	0.027	0.035	0.035
	(0.029)	(0.027)	(0.029)	(0.029)
No. of obs.	39,263	39,665	37,958	37,958

Panel A presents the treatment effects on management practices (measured using the DWMS). The outcome in Column 1 is the composite index of management practices, while Columns 2–5 display the outcomes for individual components of the management index. Finally, Column 6 has the additional dimension, leadership; the SEP asked for this dimension to be measure in addition to the four traditional components of the DWMS. The overall DMWS index used in Column 1 excludes the leadership dimension to ensure comparability with other settings. Panel B presents the treatment effects on learning outcomes (measured using PLANEA scores). The outcomes are math test scores (Column 1), language test scores (Column 2), the average across subjects (Column 3), and a composite index across subjects (Column 4). All regressions account for the randomization design (i.e., they include strata fixed effects). Panel A regressions also include enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the direct training intervention improved management practices relative to the indirect training (see Panel A, Table 2.4). Management scores in schools that received direct training were 0.13σ (p-value 0.018) higher than in "train the trainer" schools. Therefore, our results show that it pays off to invest in professional trainers to improve school principals' management capacities.[20]

Given the nature of the intervention (direct vs. indirect training on the Stallings and the SisAT tools) it is not surprising that the "Operations" and "Monitoring" dimensions improve the most. "Operations" partially measures whether there is data-driven planning, as well as personalization of instruction and learning — goals the Stallings and the SisAT specifically help with. Likewise, "Monitoring" partially measures whether school performance is measured frequently and appropriately (SisAT does this for students, and Stallings does it for teachers). Given the limitations principals face to dismiss or promote teachers, it is not surprising that the treatment effect on "People/talent management" is lower. However, measuring teachers' performance (via Stallings) enables principals to provide soft incentives (e.g., better teaching assignments or non-pecuniary rewards).[21]

While management practices improved as a result of the direct training intervention, test scores did

[20] According to surveys administered to principals as part of the nationwide student standardized test (PLANEA-Contexto surveys), in 2018 "direct training" principals were not more likely than those trained using the cascade method to undertake activities to improve learning outcomes, observe classroom teaching, help teachers improve Table A.4). However, these self-reported measures are likely inflated by social desirability bias given the (likely unrealistic) high proportion of principals who report doing these activities often or very often. Thus, we do not believe the difference between the "train the trainer" and "direct training" from these self-reported measures accurately reflects treatment effects.

[21] We further explore whether it is reasonable to expect that providing training on two tools would improve managerial practices in Section A.2. We address this question by looking at the correlation between the self-reported information on the use of the Stallings classroom observation and SisAT tools on both DWMS. We find that "direct training" schools are more likely to use the management tools provided to them, and the use of these tools is correlated with the DWMS. However, since schools that answered these surveys are statistically different from those that did not in several observable characteristics, including treatment status (see Tables A.9–A.12), these correlations may be biased and are presented for completeness.

not (see Panel B, Table 2.4). Students in "direct training" schools scored 0.03σ (p-value 0.24) higher than those in "train the trainer" schools. We can rule out, at the 95% confidence level, the possibility that test scores increased by more than 0.09σ with respect to "train the trainer" schools. This result is robust to a series of student- and school-level controls (see Table A.6). Including controls allows us to rule out an effect greater than 0.08σ at the 95% level. Finally, there is no evidence that the "direct training" affected other outcomes such as grade repetition or enrollment rates (see Table A.8).

Discussion: The lack of effect of direct training on test scores

As mentioned above, Bloom et al. (2015) find that a one-standard-deviation increase in the WMS index is associated with an increase in pupil outcomes of 0.2σ– 0.4σ. The evidence from our baseline shows a weaker correlation between management practices and test scores. Thus, optimistically assuming that a one-standard-deviation increase in management practices generates a treatment effect of 0.4σ on student learning, an increase of 0.13σ in management practices should yield an increase in test scores of 0.029σ—the actual treatment effect was 0.03σ.

We also estimate the effect of an increase in the DWMS index on test scores using the treatment assignment to instrument for the DWMS index. While this requires a strong assumption that the DWMS completely captures any possible effect of the treatment on test scores, it provides a different benchmark of the plausible causal effect of improvements in management practices on test scores. The instrumental variable approach suggests increasing the DWMS by one standard deviation increases test scores by 0.49σ (see Table A.7). This implies an expected increase of .065σ in test scores, given the treatment effect on DWMS scores.

Further, the components of the DWMS index that Bloom et al. (2015) find are more associated with test scores, are the ones where the direct training intervention improved management practices the least relative to the indirect training (see Columns 2-4 in Panel A of Table 2.4). Specifically, the treatment effect on the two components that have the highest association with learning outcomes

(“People/latent management” and “Target setting”) are the lowest.

Overall, the expected treatment effects on learning outcomes (given the treatment effects on management practices) are of the same order of magnitude as the actual treatment effects. While the direct training intervention improved management practices relative to the indirect training, these improvements did not generate statistically significant changes in learning outcomes (even with a sample size of 1,198 schools). However, we cannot rule out the possibility that management had a small positive impact on learning.

Given the low overall attendance rate to the training workshops, we explore whether increasing participation in the training workshops would result in further improvements in management practices and larger learning gains. To answer this question, we use an instrumental variable approach to study the effects of attending more training workshops. Specifically, we instrument attendance to training workshops with whether a school was randomly assigned to “direct training”. However, we face a trade-off between two different approaches to measure workshop attendance. We could use PLANEA-Contexto surveys, which all principals answered, but that do not ask about training workshops from our program specifically, but rather about any courses or counseling on how to carry out school director duties in the past. On the other hand, using the online surveys to measure (self-reported) attendance to the training workshops in this program will likely induce sample selection bias since the characteristics of schools that answered the survey are different from those that did not. We report both. While neither approach is perfect, both suggest similar results.

Using the PLANEA-Contexto surveys suggests that attending any courses or counseling on how to carry out school director duties increases both management practices and learning outcomes (see Panel B, Table 2.5). The local average treatment effects (LATE) here represent the effects of attending any workshops, not just those related to our program, for the compilers who are more likely to attend a workshop due to the “direct training” treatment. While attending any courses or counseling on how to carry out school director duties is likely to capture a significant portion of the effect of "direct training", it is unlikely to be the only channel through which the treatment affects outcomes — a necessary condition for the LATE to be valid.

Using the online surveys suggests that attending the training workshops from this program increases both management practices and learning outcomes (see Panel A, Table 2.5). However, the local average treatment effects (LATE) are likely biased due to sample selection caused by the differential attrition in the survey. In addition, and as mentioned above, the training hours across the treatment arms (cascade vs. direct) are not directly comparable.

Overall, while both approaches have limitations, they suggest one way to boost the intervention's impact on management practices and learning outcomes would be to increase principals' attendance to the training workshops.

Table 2.5: Effects of principal's attendance to the training workshops

	(1) DWMS	(2) DWMS	(3) PLANEA	(4) PLANEA
Panel A: Online surveys				
Attended > 10 hrs of training	0.36***		0.15*	
	(0.11)		(0.070)	
Attended all trainings		0.69**		0.28*
		(0.21)		(0.13)
N. of obs.	808	808	28,906	28,906
F test (first stage)	292	143	240	138
Panel B: PLANEA - Contexto				
Ever	1.1*		.68*	
	(.55)		(.39)	
Past 12 months		1**		.56*
		(.5)		(.31)
N. of obs.	850	850	29,731	29,731
F test (first stage)	16	26	13	30

Panel A presents the effect of a principal attending at least 10 hours of training on the DWMS score (Columns 1) and the overall PLANEA score (Column 3), as well as the effect of a principal attending all training on the DWMS score (Columns 2) and the overall PLANEA score (Column 4). Attendance (in both cases) is instrumented with the treatment allocation. The F statistic of the first stage is presented in the bottom row (see Table 2.2 for details on the first stage). Columns 1–2 use data at the school level, while Columns 3–4 use data at the student level. Attendance is measured using online surveys which have differential attrition across treatments (see Tables A.9–A.12). Panel B presents the effect of a principal ever attending a training workshop (on any topic related to his or her duties) on the DWMS score (Columns 1) and the overall PLANEA score (Column 3), as well as the effect of a principal attending a training workshop (on any topic related to his or her duties) in the past 12 months on the DWMS score (Columns 2) and the overall PLANEA score (Column 4). Attendance (in both cases) is instrumented with the treatment allocation. The F statistic of the first stage is presented in the bottom row (see Table A.4 for details on the first stage). Columns 1–2 use data at the school level, while Columns 3–4 use data at the student level. Attendance is measured using PLANEA-Contexto surveys which do not have differential attrition across treatments (see Table 2.1). All regressions account for the randomization design (i.e., they include strata fixed effects) and include enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, $^{**}p < 0.05$, *** $p < 0.01$

Heterogeneity

This section explores heterogeneous treatment effects on management practices by schools' (and principals') baseline characteristics. Overall, there is little evidence of heterogeneity. Specifically, we estimate the following equation:

$$Y_{isg} \quad = \quad \alpha_g + \beta_1 treatment_s + \beta_2 treatment_s \times c_s + \beta_3 c_s + \varepsilon_{isg} \qquad [2.2]$$

where c_s denotes the school characteristics of which we wish to measure heterogeneity, and β_2 allows us to test whether there is any differential treatment effect. Everything else is as in Equation 2.1. We study heterogeneity in schools' baseline management quality, marginalization index, and principals' gender and tenure. Overall, we find no evidence of heterogeneity in management practices (DWMS) or learning outcomes (see Tables A.13 and A.14).

We also study whether there is heterogeneity by whether there was a change in the school's principal between 2015 and 2018. We first assess that the treatment did not have an impact on principal turnover itself (see Table A.15) but note that 43% of schools change principals at some point in those three years. While high principal turnover may be a barrier to improving learning outcomes (Miller et al., 2019), there is no heterogeneity in treatment effects on management practices or learning outcomes by teacher turnover (see Table A.16).

2.5 Conclusions

Recent studies have identified the pivotal role that managerial practices play in helping an organization to achieve its objectives (Bender et al., 2018), and the education sector is no exception. This paper reports some of the first experimental evidence of the relative effectiveness of two interventions to improve school management in a developing country. We randomly assigned a group of public primary schools in seven Mexican states to receive training either directly from professional trainers or a "train the trainer" cascade model. Compared to indirect training, direct training improved school principals’ managerial capacity but failed to improve learning outcomes significantly. To improve student learning in the short term, a management intervention may need to have a greater impact on school principals’ managerial capacities. However, given the cost of the “direct training” intervention (~470 USD per school), the marginal dollar in Mexico might be better spent on interventions that focus on improving pedagogy (e.g., teaching at the right level, teacher content and pedagogical training) and improving teacher accountability (Kremer, Brannen, & Glennerster, 2013; Glewwe & Muralidharan, 2016; Snilstveit et al., 2016.

Chapter 3. Rule–based civil service: evidence from a nationwide teacher reform in Mexico[£]

3.1 Introduction

An efficient bureaucracy is increasingly recognized as a cornerstone of state capacity and economic development. However, historical instances indicate that putting in place that cornerstone often resembles a walk through a long road (Grindle, 2012). For example, there is consensus that the meritocratic recruitment of public officials is a key feature of a well- functioning bureaucracy (going back to the Weberian definition of a professional bureaucracy). Yet, international indexes of bureaucratic quality indicate that many governments in low—and middle—income countries fail to do so (Besley et al., 2021). This calls for a better understanding of the mechanisms used to recruit civil servants and the factors shaping their effectiveness.

In this paper, we study the effect of a civil service reform on the skills profile of new public sector employees. We do so in the context of a nationwide education reform adopted in Mexico in 2014, which revamped the teachers' civil service (the SPD reform in the rest of the paper). Among other changes, the SPD reform mandated, centrally managed, competitive examinations to determine hire and promotion decisions, proscribing a discretionary system in which such decisions were taken at the state level by local officials and teacher union representatives. The SPD reform scaled-up a previous reform which introduced the use of competitive examinations for teacher hiring in response to multiple criticisms about opacity, corruption, and absence of merit in the discretionary system.

Teachers are among the most important public service providers. They are a key input of the quality of educational services and therefore critical to productivity and long-term growth. At the individual

[£] This chapter has as coauthors Rafael de Hoyos (World Bank) and Ricardo Estrada (CAF).

level, the positive impact of being assigned to a highly effective teacher persists into adult life (Rivkin et al., 2005; Chetty et al., 2014). Furthermore, teachers tend to represent a significant part of public sector employment. In Mexico, around 24% of public sector employees are teachers—according to our own calculations using data from the national labor force survey for 2014.

We focus the analysis on the effects of the SPD reform on newly hired teachers' cognitive skills. Teacher quality is multidimensional and hard to capture in a single measure. However, several papers have shown that teachers with higher levels of cognitive skills tend to be more effective teachers—i.e., contributing with a higher value-added to student learning (Rockoff et al., 2011; Gronqvist and Vlachos, 2016; Hanushek et al., 2019; Neilson et al., 2019). Of particular relevance to contexts like the one under the study is the possibility that teaching effectively requires a basic mastery of cognitive skills. Using data from an international survey of the adult population, Estrada and Lombardi (2020) show that many teachers in Latin America, including Mexico, have low levels of literacy and numeracy, an important set of skills for effective teaching.

For the empirical analysis, we use personnel data to construct a dataset of primary and lower secondary public-school teachers hired in Mexico between 2012 and 2017. We link this data to results from the competitive examinations used to select new teachers during the same period of time. In this way, we are able to identify teachers hired through the discretionary process, before and after the SPD reform, plus successful and unsuccessful candidates to the competitive examinations. Our measure of teachers' (and applicants') cognitive skills comes from their results in "ENLACE Media Superior", a census-based national standardized test applied at the end of secondary school (grade 12) between 2008 and 2013. We focus the analysis on recent university graduates who entered (or applied to enter) the public education system between 2012 and 2017.

Results show that teachers hired after the SPD reform have higher levels of cognitive skills than teachers hired prior to the reform. This change is driven by an improvement in the bottom of the skills distribution of newly hired teachers. Quantile regressions estimates show that the ENLACE score of new teachers at the 0.1 quantile increased between 4.7 and 7.0 percentile points after the SPD reform (in 2012, teachers at the 0.1 quantile had average scores at the 24 percentile of the overall ENLACE grade 12 distribution). The corresponding effect for the 0.9 quantile is 1 percentile point and not statistically significant.[22] Our analysis rules out the possibility that the documented results are explained by secular trends or shocks, orthogonal to the reform, to the profile of individuals who self-select into the teaching profession.

We study several potential mechanisms to understand the forces behind these changes. First, we document that the SPD reform increased the share of teachers hired via competitive examinations from 63.6% in 2012 to 76.8% in 2014 and to, a maximum of, 86.4% in 2016. However, it did not completely eliminate discretionary hiring, which highlights the implementation challenges present in contexts with weak state capacity. Second, we show that prior to the SPD reform there was a skills gap favoring rule-based over discretionary hires, and this gap increased after the SPD reform. The enlargement of the skills gap is explained by a sharp increase in the skills of rule-based hires in the year in which the SPD reform was implemented and a progressive decline in the skills of the discretionary hires. Furthermore, our results show that the improvement in the skills profile of rule-based hires is driven by an improvement in the screening of applicants, with only a modest contribution from the self-selection channel. Given the promise of making merit the key criterion of personnel decisions throughout the teaching career, the SPD reform could have attracted higher skilled individuals into teaching. However, we only find small gains in the skills profile of rule-based applicants (of up to 1.5 percentile points at the 0.1 quantile). In contrast, we find that the SPD reform

[22] Prior to the SPD reform, incoming teachers at the 0.9 quantile already had high levels of skills, placing them at the 94 percentile of the overall ENLACE distribution.

made significantly steeper the relationship between the probability of being hired and skills. Prior to the SPD reform, a one-percentile-point increase in the ENLACE Grade 12 score was associated with an increase of 0.13 percentage points in the probability of being hired, while during the SPD regime the corresponding increase was 0.43 percentage points.

Summing up, we find that the implementation of the SPD reform improved the skills profile of new teachers. This progress was mainly driven by changes in the bottom of the skills distribution of new teachers. We propose two main channels to explain this result. First, the reform decreased the prevalence of discretionary hires, which were drawn disproportionately from the bottom of the skills distribution. Second, the reform improved the screening efficiency of rule-based hiring, making cognitive skills a more important determinant of hiring outcomes.

This paper contributes to several literatures at the intersection of personnel, education, and public economics. We contribute first to the literature on personnel economics of the state and more specifically to a recent strand of papers that study the effect of rule-based and discretionary hiring on the profile of civil servants. (Estrada, 2019; Neilson, Gallegos and Calle, 2019; Brassiolo, Estrada and Fajardo, 2020; Colonnelli, Prem and Teso, 2020; Dahis, Schiavon and Scot, 2020; Brassiolo, et al., 2021b; Munoz and Prem, 2021).[23] The paper also contributes to the growing literature showing the importance of selecting and promoting teachers in a transparent way as a critical element of improving the quality of teaching. The paper also contributes to the literature on teacher hiring in developing countries (Estrada, 2019; Neilson, Gallegos and Calle, 2019; Araujo, Heineck and Cruz-Aguayo, 2020; Brutti and Torres, 2021). Our findings complement those by Estrada (2019), who—studying the period previous to the SPD reform—finds that teachers hired under the rule-based mechanism were significantly more effective at increasing student learning than discretionary-hired

[23] In related work, Xu (2018); Voth and Xu (2020) study the effect of using discretion in promotion decisions in the context of the British Empire.

teachers.[24] Relative to this paper, our contribution is to document the selection patterns (self-selection and screening) shaping the skills profile of rule-based hires.

Our paper is also connected to the literature on civil service reforms, which is mainly focused on the U.S. context (Rauch, 1995; Ujhelyi, 2014; Ornaghi, 2019; Moreira and Pérez, 2021). We contribute to this literature by studying a large civil service reform in a different setting. Finally, we contribute to the literature on decentralization and particularly to the set of studies which highlight that the weak capacity of some local governments can compromise the success of decentralization reforms (see reviews in Bardhan (2002); Mookherjee (2015)). Our findings point out to one factor that could contribute to such weakness: a less meritocratic selection of bureaucrats.[25]

3.2 Institutional context

In 2013, the Mexican Congress enacted a major education reform which included the revamping of the teachers' civil service system. The newly created teachers' civil service (named "Servicio Profesional Docente" or SPD) mandated the use of competitive examinations to determine the hiring and promotion of teachers in preschool, primary, and secondary education levels.[26] It also implemented mandatory and periodical evaluations for in-service teachers, which could have different consequences based on their results. Good performers were entitled to receive monetary bonuses, while poor performers had to undergo training and could eventually—after repeated failings—be removed from teaching and placed in an administrative position. A detailed description of the reform and the political context is available in Islas, Calef and Aparicio (2021).

The SPD reform expanded in scale and scope the policies implemented by a 2008 reform named the

[24] Brutti and Torres (2021) also study the effect of rule-based (vs. discretionary) hiring on new teachers' effectiveness, in the context of a similar reform in Colombia.
[25] In the same vein, Brassiolo, Estrada and Fajardo (2021a) document that the political cycle has a large disruptive effect on the bureaucracies of municipal governments in Brazil (i.e., employee turnover increases after an election), but not on those of higher levels of government.
[26] The SPD reform did not cover either private schools or upper secondary schools operated by public universities.

“Alianza por la Calidad de la Educación” or ACE, which introduced the use of national competitive examinations to hire teachers for public schools (see details about the ACE reform in Estrada (2019)). As it was the case with the SPD reform, the ACE used test scores from national standardized examinations organized by the Federal authorities as the main criteria to hire teachers. However, there were some important differences between ACE and SPD. The use of competitive examinations in ACE was not mandatory. The decision of what share of vacancies would be filled by rule-based hiring was jointly taken by state authorities and the teacher’ union, and the rest were assigned at the discretion of the teachers’ union. In contrast to the SPD, the ACE reform did not introduce changes in other dimensions of the teaching career, it was limited to the selection of new teachers. Table 3.1 summarizes the main characteristics of rule-based hiring under both reforms.

Table 3.1: Teacher hiring reforms: ACE and SPD characteristics

Characteristics		ACE	SPD
Institutions in charge of competitive examinations	Regulations and oversight	National Committee with representatives from the Federal Ministry of Education and teachers' union (Comisión Nacional Rectora).	National Institute for the Evaluation of Education and Federal Ministry of Education.
	Implementation	State committees with representatives from State Ministry of Education and teachers' union (Comités Estatales de Seguimiento).	Federal and State ministries of education.
Exam's content(*)		1. Course content; 2. Teaching skills; 3. Intellectual skills; 4. Regulations, ethics and school management.	1. Course content; 2. Teaching skills; 3. Intellectual skills; 4. Regulations, ethics and school management.
Exam's length (number of questions)		80	175-240(**)
Hiring criteria		Weighted average of test score in standardized entry exam and (optionally, if chosen by the State Ministry of Education) bachelor's GPA.	Test exam Score in Standardized entry
Assignment of teachers to schools		Responsibility of state committees with representatives from State Ministry of Education and teachers' union.	Applicants can choose among available vacancies in order of their ranking established for hiring decisions.
Eligibility criteria		Bachelor degree from authorized public and private teacher training institutions. Specialization varies according to the position.	Bachelor's degree from any higher education institution. Specialization varies according to the position.

Notes: (*) In both cases, exams were specific for each type of position (primary teacher, secondary math teacher, etc). ACE comprised only one exam, while SPD was divided into two exams (only in the case of teachers applying to fill in a vacancy for indigenous languages took a third exam). (**) According to the 2014-2015 Exam Study Guide, the total number of questions amounted to 175. Subsequent Guides indicated a total of 240 questions.

Both the ACE and SPD reforms aimed at putting an end to an era in which personnel decisions in the public education system were characterized by a high use of discretion and opacity, and a prominent role played by the national teachers' union.[27] In Mexico, state governments are responsible for the operation of preschool, primary, and lower secondary schools, while the Federal government defines

[27] SNTE for its acronym in Spanish, see more about SNTE in Estrada (2019)

the national curriculum, monitors the system's performance, among other regulations, and finances a large share of the public education budget. State education ministries and the teachers' union are responsible for the selection and promotion of teachers. Prior to the SPD reform, the union was entitled to select a share of new hires in every state, while promotion decisions were made by joint-commissions with representatives from both state authorities and the union. In practice, the union had close to full control in the hiring and promotion decisions in many States, though, for example, the appointment of state officials who were also union leaders and the inclusion of participation in union's activities as a requirement for promotions (Santibanez, 2008). The system was widely criticized for its opacity and lack of merit in personnel decisions, including the practice of entitlement (which allowed teachers who were retiring to pass on their job to a relative) and the selling or renting of teaching positions.

Both ACE and SPD faced fierce opposition from the teachers' union. Despite this and other implementation challenges, the Federal Ministry of Education and the National Institute for the Evaluation of Education, an agency that gained full autonomy as part of the SPD reform, implemented in July of 2014 the first SPD examination for teacher selection, a process which took place every year until 2018. In 2019, the SPD reform was canceled, as a result of a change in the political party in office. Figure B.1 in the annexes illustrates in a timeline the significant changes in teacher personnel policies that the Mexican education system experienced throughout the 2007-2019 period.

3.3 Data

Data sources

We obtain information on primary and lower secondary school teachers hired from 2012 to 2017 from the following three databases:

- *Registro Nacional de Maestros, RENAME*: quarterly administrative dataset on school personnel (3rd quarter of 2011–4th quarter of 2012). Identifies teachers hired in 2012.

- *Censo de Escuelas, Maestros y Alumnos de Educación Básica, CEMABE*: census of schools, teachers and students carried out by the National Statistics Office, INEGI (September–December 2013). Identifies teachers hired in 2013.
- *Fondo de Aportaciones para la Nómina Educativa, FONE*: quarterly administrative dataset on school personnel (1st quarter of 2015 – 2nd quarter of 2018). Identifies teachers hired from 2014 to 2017.

RENAME and FONE are assembled by the Federal Ministry of Education and cover all public-school teachers paid with federal funds, with the exception of those based in Mexico City schools.[28] There are no official documents reporting the total number of teachers in the payroll of State governments, but according to our own estimations they accounted for around 13% to 15% of the total number of teachers hired in 2013. Most teachers held positions funded by the Federal government and are hence part of our analysis.

Data on the individuals who applied to the rule-based hiring process is based on microdata from ACE for the period 2012-13 and SPD for the period 2014-17. There are not comparable records of "applicants" to the discretionary process. To the best of our knowledge, there is no dataset documenting the process and criteria used to select and promote teachers under the discretion of the teachers' union with the tacit or explicit approval of State governments.

Our measure of teachers' (and rule-based applicants') cognitive skills comes from their results in ENLACE Media Superior (ENLACE hereafter), a census-based standardized test applied to all students in their final year of upper secondary school (grade 12) between 2008 and 2013. The exam

[28] CEMABE was designed to cover all public-school teachers and hence is inclusive of both teachers paid with federal and state funds. To avoid counting state-funded teachers as federally-teachers we exclude all teachers in CEMABE who are based in schools which do not appear in either RENAME or FONE records).

measured student achievement in mathematics, literacy, and a rotating subject. The main purpose of the assessment was to provide feedback to the different stakeholders in the education system and its results did not have consequences for students' graduation or admissions into the next schooling level. Participation in ENLACE was optional, but more than 90 percent of all students sat the test each year.[29]

Merging and Variables Included in the Final Dataset

We track individuals through the RENAME, CEMABE, and FONE data sets using their population id (CURP for its acronym in Spanish) and taxpayer id (RFC) and construct a data set of new teachers inclusive of name, gender, birth date, type of teaching position, assigned school(s), and year of hiring. We identify as new teachers those individuals who do not appear in any record from previous school years—the datasets do not include information on hiring dates.[30] We merge the applicants' data from the ACE and SPD examinations and build a dataset that contains applicants' population id, their results in the entry examination, and the type of teaching position they applied to.[31]

The results from the ACE and SPD entry examinations are merged to the teachers' data using the taxpayer number (RFC), and population id (CURP). The data does not include hiring offers nor acceptances. Hence, we define as rule-based hires those individuals who apply to the rule-based process, either the ACE or SPD examinations, and are part of the teacher payroll the following school year. Newly hired teachers that did not participate in a rule-based process and are part of the payroll are labeled as "discretionary hires".

[29] de Hoyos, Estrada and Vargas (2021) find that ENLACE test scores are a strong predictor of schooling and labor market trajectories and conclude that ENLACE was effective at capturing the cognitive skills which it was designed to measure.

[30] As mentioned above, we exclude individuals based in schools which are only reported in CEMABE, but not in RENAME and FONE as they are likely state-funded teachers.

[31] As test scores under SPD are only available in brackets according to 5 performance categories, we construct a corresponding variable for the ACE test scores–reported as a continuous variable–that mimics the distribution of SPD applicants among these categories.

The individual ENLACE test scores were transformed into year-specific percentiles of the overall distribution of the simple mean score in math and literacy. Among the 4.9 million students who sat the test between 2008 and 2012, 95 percent have a complete population id (CURP). We use the CURP to merge the ENLACE data with the teachers and applicants databases described above. The final dataset includes information on all new hires, rule-based and discretionary, for primary and lower-secondary public schools between 2012 and 2017, a period covering before and after the Reform was implemented.

Samples used in the analysis

We identify a total of 181,590 individuals hired as primary and lower-secondary school teachers during the 2012-2017 period. Because of their age, many of these individuals finished secondary school before ENLACE grade 12 was carried out for the first time in 2008. For the 2012 cohort of new teachers, the first in our analysis, we are only able to identify the ENLACE score of those hired 4 years after concluding secondary school.[32] Hence, our analysis is based on the sample of 24,914 individuals who were hired 4 years after finishing upper secondary, we refer to these individuals as "recent graduates".[33] As shown in Panel A of Table 3.2, among all new teachers in our sample—column (1)— 65% are female, they are on average 29.5 years old, and 46% were hired in a rule-based process (either before or after the SPD reform). Recent university graduates in our main sample—column (2)—are more likely to be female (69%), younger (they are 22.2 years old on average) and hired through a rule-based process (78%). Their ENLACE score is, on average, at the 67 percentile of the national distribution.

[32] A teaching degree in Mexico typically requires 4 years of studies.

[33] We use the sample of teachers hired 4 or 5 years after finishing upper secondary for robustness checks (using 2013 as baseline).

Table 3.2: New teachers and rule-based applicants: Descriptive statistics

Panel A - New Teachers 2012-2017

		All	Recent graduates (4 years)	Recent graduates (4 & 5 years)
		(1)	(2)	(3)
Female	Mean	0.649	0.691	0.686
	SD	0.477	0.462	0.464
Age	Mean	29.529	22.283	22.664
	SD	8.864	0.749	0.918
ENLACE Score (in percentiles)	Mean		67.342	64.508
	SD		25.012	25.634
Ruled-based hire	Mean	0.457	0.784	0.732
	SD	0.498	0.411	0.443
	Observations	181,590	24,914	39,615

Panel B - Ruled-based applicants 2013-2017

		All	Recent graduates (4 years)	Recent graduates (4 & 5 years)
		(1)	(2)	(3)
Female	Mean	0.692	0.764	0.752
	SD	0.461	0.424	0.432
Age	Mean	29.147	22.294	22.611
	SD	7.138	0.766	0.927
ENLACE Score (in percentiles)	Mean		61.547	59.594
	SD		26.570	26.834
Top-quartile score in entry examination	Mean	0.241	0.377	0.346
	SD	0.415	0.484	0.476
Top-half score in entry examination	Mean	0.505	0.659	0.625
	SD	0.498	0.474	0.484
Ruled-based hire	Mean	0.136	0.300	0.282
	SD	0.343	0.458	0.445
	Observations	468,846	55,660	79,530

Notes: The Panel A shows summary statistics for new teachers of the period 2012-2017. Column (1) shows statistics for all new teachers including those we cannot merge with ENLACE. Column (2) shows statistics for recent graduates who entered the system 4 years after graduating from secondary school. Column (3) shows statistics for recent graduates who entered the system 4 or 5 years after graduating from secondary school. The Panel B shows summary statistics for ruled-based applicants for the period 2013-2017. Column (1) shows statistics for all ruled-based applicants including those we cannot merge with ENLACE. Column (2) shows statistics for recent graduates who approached the ruled-based system 4 years after graduating from secondary school. Column (3) shows statistics for recent graduates who approached the ruled-based system 4 or 5 years after graduating from secondary school.

Panel B in Table 3.2 shows some descriptive statistics for the 468,846 individuals who applied to the rule-based process during the period 2013–2017. 69% of them are female, they are on average 29 years old, and 14% succeeded in being hired. There are 55,660 recent university graduates. 76% are female and they are 22 years old on average. Recent graduates tended to perform well at the standardized test used in the competitive examinations. 38% (66%) scored at the top quartile (half) of their entry examination test and 30% of them were hired. They have an average ENLACE score at the 61.5 percentile of the national distribution.

3.4 Empirical strategy

We are interested in identifying the effect of the SPD reform on the skills profile of new teachers and in shedding light on its underlying mechanisms. The SPD reform had a national reach and was implemented simultaneously across the country; therefore, we rely on time variation for our first set of results:

$$\boldsymbol{Y_{it}} = \boldsymbol{\alpha} + \sum_{\boldsymbol{\pi}=\mathbf{2013}}^{\mathbf{2017}} \boldsymbol{\beta_{\pi}} \cdot \mathbf{1}[\boldsymbol{\pi} = \boldsymbol{t}] + \boldsymbol{\Gamma} \boldsymbol{X_{st}} + \boldsymbol{\theta_s} + \boldsymbol{\epsilon_{it}} \qquad [\mathbf{3.1}]$$

Where $\boldsymbol{Y_{it}}$ is the ENLACE grade 12 test score (in percentiles) of individual *i* hired at year *t*. $\mathbf{1}[\boldsymbol{\pi} = \boldsymbol{t}]$ is a vector of dummy variables that indicate the relative time (in years) with respect to 2012, the year prior to the enactment reform, and $\boldsymbol{\beta_{\pi}}$ are the coefficients of interest. $\boldsymbol{X_{st}}$ is a vector of labor market variables as controls and $\boldsymbol{\Gamma}$ is the associated vector of parameters. $\boldsymbol{\theta_s}$ is a vector of state fixed effects and $\boldsymbol{\epsilon_{it}}$ is an error term. Heteroskedasticity robust standard errors are presented.

There was almost a one-year gap between the announcement and implementation of the SPD reform which could have led to anticipation effects. We deal with such a possibility by using 2012 as the baseline year, one year before the reform was announced and two years before it was implemented). To explore heterogeneous effects along the distribution of $\boldsymbol{Y_{it}}$, we complement the main specification (equation 3.1) with a quantile regression model:

$$\boldsymbol{Q_{\tau}(Y_{it}) = \alpha + \sum_{\pi=2013}^{2017} \beta_{\pi}(\tau) \cdot 1[\pi = t] + \Gamma(\tau) X_{st} + \theta_s + u_{it}} \qquad [\mathbf{3.2}]$$

Where $Q_{\tau}(Y_{it})$ is the ENLACE Grade 12 test score of individual *i* hired at year *t* at the *τth* quantile and $\beta_{\pi}(\tau)$ and $\Gamma(\tau)$ are vectors of coefficients at the various quantiles. For simplicity, we present results for the 0.1 and 0.9 quantiles in the main text and for the 0.25 and 0.75 quantiles in the Appendix B.

In equations 3.1 and 3.2, the effect of the SPD reform is identified using time variation and, hence, it faces two important threats to identification. First, the effect of the reform might be confounded with a secular trend, unrelated to the SPD reform, on the ENLACE scores of new teachers. To mitigate this concern, we use a dynamic specification of treatment effects— instead of a simple before and after specification—which allows to explore in a desegregated way the evolution of the new teachers' profile. Furthermore, by using the year-specific percentiles of ENLACE test scores as the outcome variable, we concentrate on changes in the relative position of newly hired teachers in the national distribution of cognitive skills, holding constant changes in overall test scores across years. We do not have information about long-term in the skills profile of the teaching force in Mexico, but a string of papers from other contexts has documented a secular decline—mostly in developed countries (Nickell and Quintini, 2002; Corcoran et al., 2004; Fredriksson and Ockert, 2008), but also in Chile in Latin America (Neilson, Gallegos and Calle, 2019)—which in our specification would lead to an underestimation of the effects of the SPD reform on the skills profile of new teachers.

A second threat to identification is related to potential shocks that affect the selection process into teaching in a contemporaneous way to the SPD reform. We do not have any information about a contemporaneous policy or economic shock that could produce such a pattern. Nonetheless, we include in the estimation of equations 3.1 and 3.2 a set of covariates on economic conditions at the local (state) level that could affect occupational decisions in the labor market: the state's GDP per capita, unemployment rate, and the mean wages earned by tertiary educated individuals working in non-teaching occupations.

3.5 Main results

Figure 3.1 shows point estimates for the change on the average ENLACE score of the teachers hired every year—coming from the estimation of equation 3.1. As said before, we focus on the hiring of recent university graduates. Confidence intervals at the 95 percent level are reported in vertical lines.[34]

We find a clear improvement in the skills profile of teachers hired after the SPD reform. The point estimates for the post-reform years vary from 2.7 to 3.8 percentiles of the ENLACE Grade 12 test score distribution and are all statistically significant at the 95 percent level. We do not find clear evidence of an anticipation effect. The coefficient for 2013 is negative, but it has a small magnitude (-1.0 percentiles) and is not statistically significant at conventional levels. The improvement in the skills profile of teachers hired after the reform is economically meaningful. In context, teachers hired in 2012—our baseline year—had on average an ENLACE score at the 65th percentile of the test

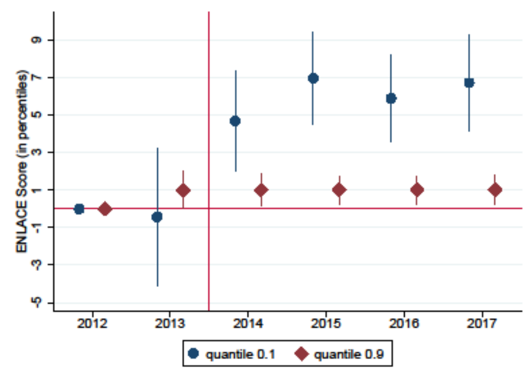
[34] Figure B.2a presents the raw yearly means in the Appendix and Table B.1 reports the point estimates plotted in Figure 1a. It also presents results from estimations without the controls for the state-level labor market conditions and using yearly changes in these variables instead of levels. Results are consistent across specifications.

distribution.

Figure 3.1: New teachers: Change in ENLACE scores with respect to 2012

(a) OLS estimates (2012: mean = 64.9)

(b) Quantile regression estimates (2012:p10 = 24.3, p90 = 94.1)



Notes: The figure (a) shows the difference in percentiles of the ENLACE Score of newly hired teachers with respect to 2012, corresponding to the $\beta\pi$ estimates of equation (3.1). The figure (b) shows the difference in percentiles of the ENLACE Score of newly hired teachers for quantile 0.1 in blue and quantile 0.9 in red with respect to 2012 corresponding to the $\beta\pi(0.9)$ and $\beta\pi(0.1)$ estimates of equation (3.2). Regressions include state fixed effects and state job market controls and robust standard errors. Confidence intervals at 95% level are shown in bars.

We estimate equation 3.2 to study whether the change in the profile of teachers hired after the reform varied across the skills distribution. Figure 3.1b reports the corresponding point estimates for the 0.1 and the 0.9 quantiles—see the plotted coefficients in Table B.2 and the raw yearly percentiles in Figure B.2b in the Appendix. The SPD reform was accompanied by a sharp improvement in the level of skills of teachers at the 0.1 quantile. The point estimates for the post-reform years vary from 4.6 to 6.2 percentiles of the ENLACE score and are all statistically significant at the 95 percent level. The magnitude of these point estimates is economically significant. In 2012, new teachers at the 0.1 quantile had ENLACE test scores at the 24 percentile of the national test distribution. On the contrary, the corresponding point estimates for the 0.9 quantile are small in magnitude (around 1 percentile) and not statistically significant. The test scores of incoming teachers at the 0.9 quantile were already high in 2012 (equivalent to the 94 percentile of the overall ENLACE distribution) so there was little scope to improve. A similar —though less contrasting—pattern is found from comparing instead the 0.25 and the 0.75 quantiles—see Figure B.2b. These findings indicate that the improvement in the skills profile of new teachers produced by the SPD reform is mainly explained by an improvement in the bottom of the skills distribution.

For data limitations, we focus our analysis on teachers who were hired four years after finishing secondary school, the so-called recent graduates. Figure B.3 in the Appendix shows that our results are robust to the inclusion of individuals who were hired five years after finishing secondary school—unfortunately, we cannot go beyond this extension— using 2013 as the baseline year.

3.6 Mechanisms

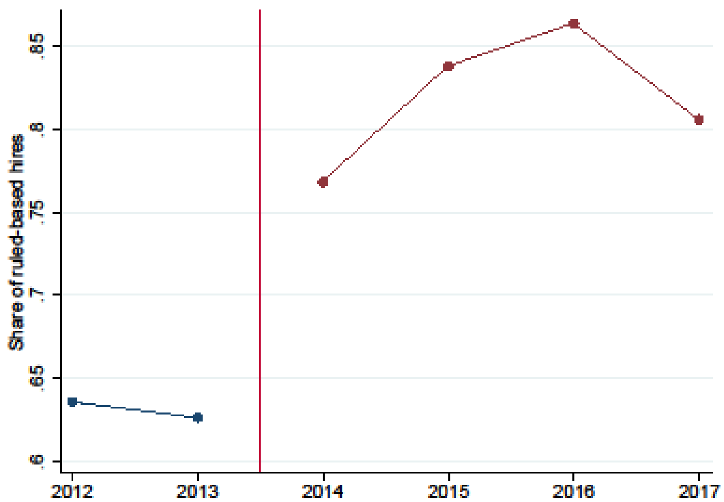
The implementation of the SPD reform was followed by an increase in the average ENLACE score of the newly hired teachers. Such a pattern suggests a positive effect of the rule-based personnel policies implemented with the reform on the skills profile of new teachers. We turn now to investigate evidence that could support such causal interpretation and shed light on the mechanisms linking the SPD reform with the improvement in teachers' skills.

The prevalence of ruled-based hiring

We study first the extent to which the SPD reform effectively increased the prevalence of rule-based hiring in the public education system. Figure 3.2 reports the share of new teachers hired under the rule-based method. First, rule-based hiring was the majoritarian way of selecting teachers even before the SPD reform, with 63.6 percent of the teachers hired through a competitive examination in 2012—as result of the ACE reform described in Section 3.2. Second, the SPD reform significantly increased the prevalence of rule-based hiring—by 14 percentage points in its first year of implementation and

23 percentage points two years later (see point estimates in Figure B.5). Third, although to a lesser extent, the use of discretionary hiring persisted after the SPD reform.[35] Since, as it will be shown in the following section, rule-based hires have higher skills vis-a-vis discretionary hires, the increase in the prevalence of rule-based hires is consistent with a positive impact of the SPD reform on the skills profile of new teachers.

Figure 3.2: New teachers: Share of ruled-based hires



Notes: The figure shows the share of ruled-based newly hired teachers for each of the sample years. As explained in Section 3.1, we define rule-based hires as those individuals who apply to the rule-based process (ACE or SPD) and are hired as teachers the following school year.

The Skills Gap Between Rule-based and Discretionary Hires

Figure 3.3a plots yearly means of ENLACE scores (in percentiles) by hiring method. Three stylized facts stand out. First, rule-based hires have on average higher ENLACE scores than discretionary hires starting from the baseline year. Second, there seems to be a discontinuity in the skills profile of

[35] We might have overestimated the proportion of rule-based hires among total hires. As we do not observe job offers, we define rule-based hires as the individuals who apply to the rule-based process and are hired as new teachers, which might include individuals who, despite having participated in a competitive examination, were hired through the discretionary process.

rule-based hires when the reform is implemented. Third, the skills profile of discretionary hires follows a downward trend. Hence, the skills gap between rule-based and discretionary hires is positive and increasing over time.

Figure 3.3b shows regression estimates for the skills gap between rule-based and discretionary hires by year of hiring, which confirm the positive and widening skills gap between these two groups of teachers. 2012 rule-based hires have on average a 3.4 percentile higher ENLACE score than discretionary hires. The gap increased to 6.7 in 2013 and to 14.5 in 2014—the first year in which the reform was implemented. As said before, the widening gap is mainly explained by a marked improvement in the level of skills of rule-based hires between 2013 and 2014 (the point estimate for the difference between these two years is 4.4 percentiles, with p-value of 0.01), and a downward trend in the level of skills of discretionary hires (the ENLACE scores of this group decreased on average by 3.2 percentiles per year from 2012 to 2016).

Figures 3.3c and 3.3d present raw yearly percentiles and quantile regression estimates for the skills gap between rule-based and discretionary hires. Results are reported for the 0.1 and 0.9 quantiles. As these figures show, the increase in the gap at the 0.1 quantile is startling. The gap moved from 5 percentiles in 2012 to 20 percentiles in 2014. In contrast, the corresponding gap at the 0.9 quantile is 2 percentiles in 2012 and 3 percentiles in 2014.[36]

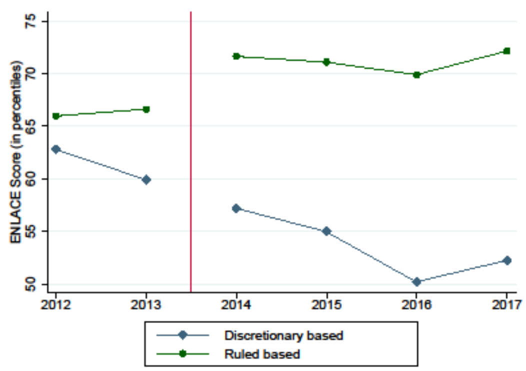
Combined with the increased prevalence of rule-based hiring, the positive skills gap between rule-based and discretionary hires and—particularly—its large increase in the year in which the SPD reform was implemented support the hypothesis that this reform is responsible for the documented improvement in the skills profile of new teachers. The sharp improvement in the skills profile of rule-

[36] See OLS and quantile regression estimates for the yearly change in the skills gap with respect to 2012 in Figures B.6a and B.6b; and raw yearly percentiles and quantile regression estimates for the 0.25 and 0.75 quantiles in Figures B.7a and B.7b.

based hires in 2014 but not in discretionary hires is evidence against the concern that the improvement in the skills profile of new teachers was the result of a secular trend in the profile of individuals who self-select into the teaching profession (which would affect both rule-based and discretionary hires), a matter which we investigate further below.

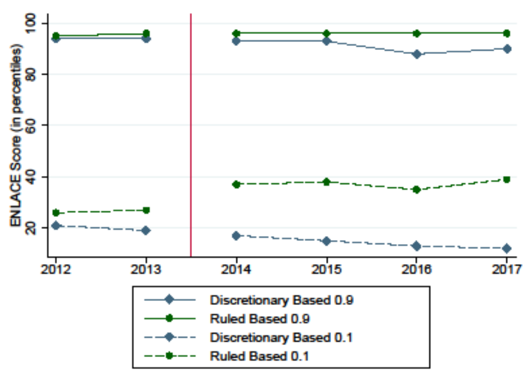
Figure 3.3: New teachers: ENLACE scores of ruled-based vs. discretionary hires

(a) Means

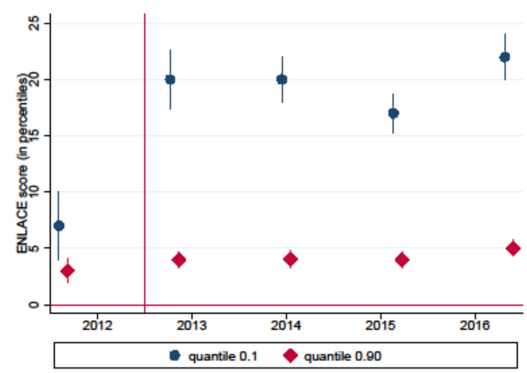


(b) OLS estimates: yearly gaps

(c) Percentiles



(d) Quantile regression estimates



Notes: The figure (a) shows the ENLACE Score in percentiles means for ruled-based and discretionary-based newly hired teachers for each year in the sample. The figure (b) shows the estimates of the annual gaps in ENLACE Score in percentiles between ruled-based and discretionary-based newly hired teachers, regressions include state fixed effects and robust standard errors. The figure (c) shows the ENLACE Score in percentiles at quantiles 0.1 and 0.9 for ruled-based and discretionary-based of newly hired teachers for each year in the sample. The figure (d) shows the estimates of the annual gaps in ENLACE Score in percentiles at quantiles 0.1 and 0.9 between ruled-based and discretionary-based newly hired teachers, regressions include state fixed effects and robust standard errors. For figures (b) and (d) confidence intervals at 95% level are shown in bars.

Self-selection and Screening in Rule-based Hiring

Two channels could explain the documented change in the skills gap between rule-based and discretionary hires: (changes in) self-selection and screening. We study the importance of both mechanisms analyzing the rule-based process; we cannot do the same for the discretionary process because we only observe those individuals who were hired, and not the full pool from which they are selected).

Self-selection

To study whether the pool of applicants to rule-based hiring changed after the SPD reform, Figure 3.4a shows point estimates for the change with respect to 2013 on the ENLACE score of rule-based applicants. There is no evidence of an improvement on the skills profile of rule-based applicants. In fact, between 2013 and 2014 there is a small drop on the average ENLACE score of rule-based applicants (point estimate is -1.4 percentiles, with p-value of 0.00), while from 2015 and on there are not statistically significant changes with respect to 2013.

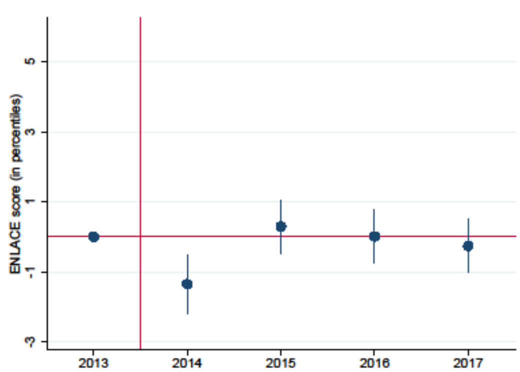
Figure 3.4b plots the corresponding quantile regression estimates. For the 0.9 quantile, there are significant decreases of around 1 percentile in 2014 and 2017 with respect to 2013, and no observed change in 2015 and 2016 with respect to the same year. For the 0.1 quantile, there is no observed change between 2013 and 2014, and then there is a significant increase of around 1.5 percentiles from 2015 and onwards. Hence, at least at the bottom of the skills distribution there was a modest improvement in the profile of those individuals who decided to self-select into the rule-based process.[37]

[37] We obtain a similar picture if instead of focusing only in individuals who finished secondary school four years prior to applying to a teaching position, we include in the sample those who finished four or five years prior, as it is shown in Figure B.8 in the Appendix.

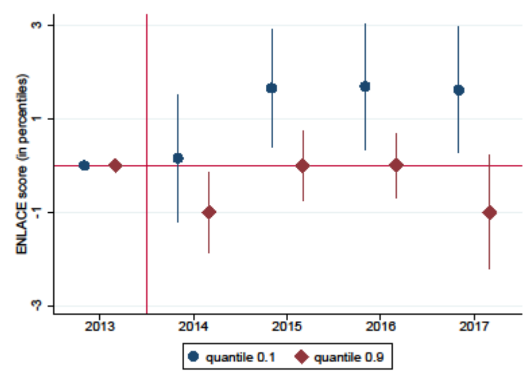
To interpret the magnitude of the documented patterns of self-selection, consider that with respect to 2013, in the 2014–2016 period the average ENLACE score of rule-based hires improved by around 3.4–4 percentiles and by 5–7 percentiles at the 0.1 quantile of the rule-based hires' distribution (see Figure B.9). This implies that the contribution of the self-selection channel to the improvement in the skills profile of rule-based hires is at most modest, which further weakens the concern that the main results could be explained by a secular trend or shock—independent of the reform—to the skills profile of individuals who self-select to teaching.

Figure 3.4: Rule-based applicants: Change in ENLACE scores with respect to 2013

(a) OLS estimates (2013: mean=59.0)



(b) Quantile regression estimates (2013: p10=19.6, p p90=92.5)

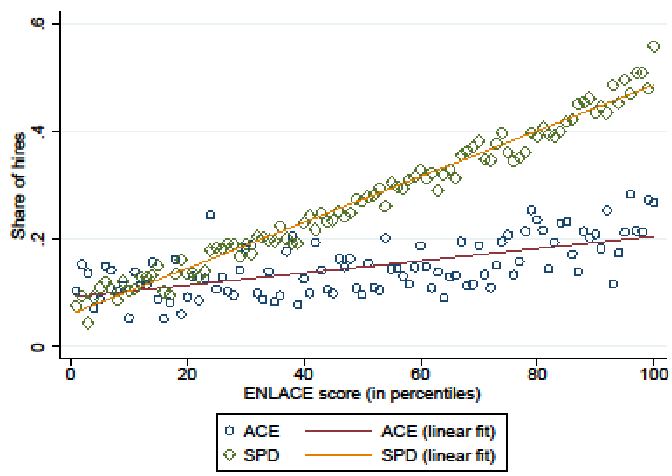


Notes: The figure (a) shows the difference in percentiles of the ENLACE Score of ruled-based applicants with respect to 2013, corresponding to the βπ estimates of equation (3.1). The figure (b) shows the difference in percentiles of the ENLACE Score of ruled-based applicants for quantile 0.1 in blue and quantile 0.9 in red with respect to 2013 corresponding to the βπ(0.9) and βπ(0.1) estimates of equation (3.2). Regressions include state fixed effects and robust standard errors. Confidence intervals at 95% level are shown in bars.

Screening

The above results suggest that with the SPD reform, the rule-based process became more effective at producing high-quality hires given a pool of applicants with a relatively constant level of skills. To investigate this, Figure 3.5 reports local means of the probability of being hired among rule-based applicants by percentile of the ENLACE score for the periods before and after the SPD reform. As expected, the relationship between the probability of being hired and the ENLACE score changed significantly with the implementation of the SPD reform, making skills a stronger predictor of hiring (Figure B.11 in the Appendix reports similar results by year to show this change is not the result of a secular trend). In the ACE regime, a one-percentile-point increase in the ENLACE score was associated with an increase of 0.13 percentage points in the probability of being hired, while during the SPD regime the corresponding increase was of 0.43 percentage points. While an applicant with an ENLACE score at the 10th percentile had the same (unconditional) probability of being hired in both the ACE and SPD examinations (of 10.9%), an applicant with an ENLACE score at the 90th percentile had a 20.9% chance of being hired during the ACE examination and a 43.5% chance during the SPD one.

Figure 3.5: Ruled-based applicants: Probability of being hired by ENLACE score



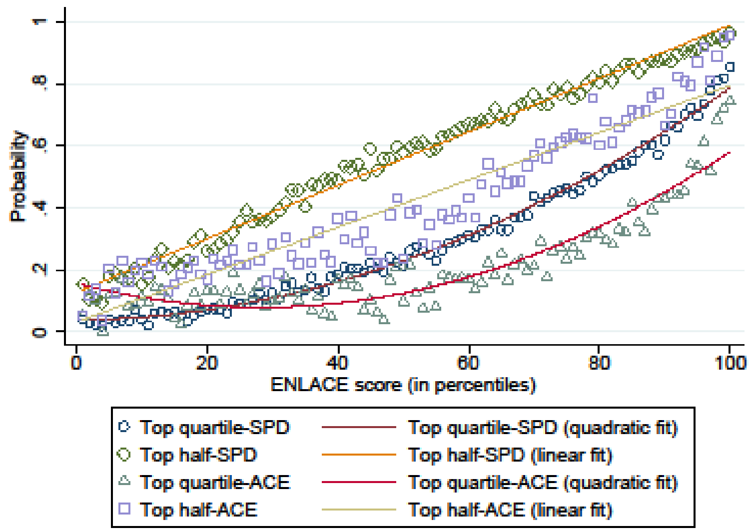
Notes: The figure shows probability of being hired by each percentile in ENLACE Score for ACE and SPD ruled-based applicants. A linear fit is included for each system.

As described before, both the SPD and ACE regimes used standardized tests to screen applicants. However, the institutions responsible for designing the recruitment processes were different, as well as the length and contents of the tests. Hence, it is possible that the screening efficiency of the SPD and ACE examinations differs.

Figure 3.6 plots local means of the probability of scoring in the top quartile(half) of the SPD(ACE) entry examination by the applicant's ENLACE score controlling for state and type of teaching position fixed effects.[38] In line with the previous results, the SPD examination strengthened the relationship between screening and cognitive skills. For example, an ACE applicant with an ENLACE score at the 50th percentile had an (unconditional) probability of scoring at the top quartile(half) of the entry examination distribution of 9.8% (23.6%), while an SPD applicant with the same ENLACE score had a probability of 21.3% (58.0%) of scoring at the top quartile(half) of the entry examination. Summing up, the SPD regime was more successful than its rule-based predecessor, ACE, in making hiring outcomes more dependent on applicants' cognitive skills.

[38] Figure B.11 reports similar plots by year.

Figure 3.6: Rule-based applicants: Probability of obtaining a high entry examination score by ENLACE Score



Notes: The figure shows the probability of scoring in the top 25% and top 53% of the ACE and SPD entrance exams for each of percentiles in ENLACE Score. For the probability of scoring in the top 53% in any of the systems a linear fit was included, in the other hand, for the probability of scoring in the top 25% in any of the systems a quadratic fit was included.

3.7 Conclusions

The results presented in this paper show that the implementation of the SPD reform improved the skills profile of new teachers. This progress was mainly driven by changes in the bottom of the skills distribution of newly hired teachers. We propose two main channels to explain this result. First, the reform decreased the prevalence of discretionary hires, which were drawn disproportionately from the bottom of the skills distribution. Second, the reform improved the screening efficiency of rule-based hiring, making cognitive skills a more important determinant of hiring outcomes.

The SPD reform decreased, but did not eradicate, the use of discretionary hiring, a finding that had not been documented before in this context and that echoes evidence from a similar teacher reform held in Colombia (Brutti and Torres, 2021). The incomplete reach of these reforms speaks about the challenges of implementing civil service reforms in contexts where institutions might be weak and incentives to non-compliance are strong. The skills profile of discretionary hires was already worse than those of rule-based hires and worsened after the SPD reform. This piece of evidence is consistent with the idea that the abuse of discretion in hiring can lead to the selection of lower skilled individuals because they are the ones who have fewer labor market options and therefore benefit the most from non-meritocratic access to a job (Estrada, 2019).

The SPD reform was based on the promise of making merit the key criterion of personnel decisions through the teaching career, in contrast to the accusations often made to the discretionary regime. One could expect that such a change in personnel policy made teaching a more attractive career for higher skilled individuals, a possibility reinforced by the fact that the reform eliminated application restrictions among university graduates from fields different that the ones exclusively related to teaching ("escuelas normales" as they are known in Mexico). We only find a modest improvement in the skills profile of rule-based applicants. However, we cannot discard that the self-selection effect would have grown in importance in the medium- to long-term (as our post-reform period of analysis only comprises 4 years).

Several countries in Latin America have implemented teacher reforms to make rule-based hiring mandatory—the list includes Brazil, Chile, Colombia, Ecuador, Mexico, and Peru. The results presented in this paper highlight two important dimensions to evaluate the effect of such type reforms on teacher quality. First, its overall efficiency vs. the alternative regime. Second, the technical efficiency of the specific test used to screen teachers.

The results presented in this paper show how a rule-based civil service system can improve the skills profile of incoming public officials. However, the ambitious reform under analysis faced a political backlash that led to its cancellation. This highlights that the political implementation and the generation of a broad coalition of support among public servants and other stakeholders is as important as the technical content of personnel policies for the long-run success of civil service reforms.

Chapter 4. Doing with one hand and undoing with the other: conflict de-escalation, illicit economies, and school dropout risk[∝]

4.1 Introduction

From the diamond fields of West Africa to the coca crops of Colombia, armed conflict financed by the illicit exploitation of natural resources has an extremely negative impact on education (Justino, 2016). It can discourage educational investment, bring about the destruction of school infrastructure, reduce teacher attendance due to fear, and produce student absenteeism derived from physical menaces and parental restrictions.[39] Therefore, while the end of an armed conflict and the de-escalation of the resulting violence may positively affect educational indicators, such conflicts deeply permeate the productive structures of the territories where it develops, creating or exploiting ways to obtain funds from illicit activities (Arjona, 2016; Santos, 2018).

Civilians living in these areas are exposed to a paradoxical reality: although they suffer the destruction, death, and fear caused by conflict, they nevertheless adapt to it, coping with the violence and often becoming involved in the illicit activities that finance the struggle. The end of the conflict does not necessarily result in a reduction of the illicit activities that contributed to fund it. In fact, if such activities continue, or even grow—as occurred with coca crops in Colombia—, they may negatively affect the educational outcomes of children and youth in those areas much more than they did when the violence was rampant. To the best of our knowledge, causal evidence of the changes in educational decisions related to the end of violence under the presence of labor opportunities

---

∝ This chapter has as coauthors Alejandra Quintana (Columbia University) and Fabio Sánchez (Universidad de los Andes).

39 A large body of literature (Justino, 2016) has documented the negative effects on educational outcomes of children exposed to violence from civil conflicts. Children and youth in conflict areas have lower educational attainment, are less likely to enroll in school, and are more likely to drop out (Akresh and de Walque, 2008; Chamarbagwala and Morán, 2011; Dabalen and Saumik, 2012; Leon, 2012; Merrouche, 2006; Parlow, 2012; Rodriguez and Sanchez, 2009; Shakya, 2011; Valent, 2013). In scenarios of armed conflict, access to public goods in general and to educational public goods in particular in the most affected areas becomes difficult and expensive (Guariso and Verpoorten, 2019; Minoiu and Shemyakina, 2014; Diwakar, 2015).

stemming from illicit activities such as coca crops is sparse. This paper contributes to filling this gap by studying the unique setting created by the peace negotiations undertaken by the Colombian government and the FARC-EP between 2012 and 2016.

In Colombia, coca cultivation has been a key illicit activity that has financed a vast array of armed groups. It has had pervasive effects on communities that live near to the coca crops, including children and youth. The existing literature has found that exposure to illegal labor markets during childhood affects children's human capital formation (Sviatschi, 2018). Also, there is evidence that booms in coca prices increase the labor supply of teenage boys (Angrist and Kugler, 2008). In 2018, two years after the peace negotiations ended, only 32 percent of the school-age population in Colombia's coca growing municipalities attended school, while 92 percent of children between the ages of 6 and 8 worked (UNODC and Fundación Ideas para la Paz, 2018). Evidence uncovered by investigative journalists suggests that Colombian children and youth leave school in most cases to harvest coca (Aleteia, 2016; *The New York Times*, 2017). However, little is known about how households in areas heavily affected by armed conflict and exposed to labor markets stemming from illicit activities modify their schooling decisions because of conflict de-escalation.

In this paper, we exploit the de-escalation of Colombia's armed conflict during the 2010s in order to assess the effect of a reduction of violence on the educational outcomes of children and youth residing in the territories highly affected by the confrontation. We ask whether the reduction of violence led to the improvement of educational outcomes, particularly dropout rates. We take advantage of the fact that in 2012 the Colombian government and the country's largest guerrilla force, the Revolutionary Armed Forces of Colombia –People's Army (henceforth FARC-EP) entered negotiations to end the decades-long civil conflict. The peace dialogues led to several political and developmental agreements, including two events that were critical for our research. First, in

December 2012, the FARC-EP declared a ceasefire as a credible sign of commitment to the peace agreement. Second, in May 2014, both parties announced that the government would provide monetary incentives for those peasants who voluntarily substituted their coca crops for other legal crops, creating opportunities for them in legal economies.

To identify the impact of the ceasefire on school dropout, we first construct the educational trajectories of children and youth in school—a variable equal to zero (0) during the years the student is enrolled in the school and equal to one (1) in the year the student drops-out. Second, we econometrically determine whether the likelihood of a dropout event diminished in the municipalities most exposed to FARC-EP violence after the unilateral ceasefire. To causally identify the effect of the ceasefire on the dropout risk in the most violent territories, we implement a difference-in-differences methodology with household fixed effects. The use of household fixed effects would correct the estimates for the likely bias stemming from the fact that parents less concerned about their children's education may choose to live in highly violent municipalities. Thus, we use information on the educational trajectories of children and youth living in households with two or more siblings.

Our work relies on the following sources of information: (a) the annual censuses of primary and secondary education students of the country (SIMAT); (b) the registry of population and households potentially eligible for social programs (SISBEN); (c) data from the National Center for Historical Memory (CNMH), which has compiled information on conflict variables and other relevant indicators; and (d) the municipal panel from the Economic Development Research Center at the University of the Andes, which gathers a variety of socioeconomic, geographic, and political variables.

Our results show that conflict de-escalation, which reduced FARC-EP violence by 90 percent, led to an increase in dropout rates. We find that the dropout risk rose by 0.9 percentage points, an increase of 13 percent relative to the mean of schools highly exposed to FARC-EP violence municipalities prior to the ceasefire, and 15 percent relative to the mean in municipalities that were not highly exposed to FARC-EP violence after the ceasefire. The results obtained reveal that the de-escalation of the armed conflict in Colombia negatively impacted school attendance of the most vulnerable students living in the municipalities affected by the actions of the FARC-EP.

The effects on primary and secondary school students are explored separately. The dropout risk for primary school students increased by 5.4 percent after the ceasefire, relative to the mean in the same municipalities prior to the ceasefire. By contrast, for secondary school students, it increases by 20 percent. Thus, the likelihood of dropping out of school is concentrated among the older students. Moreover, the effect found is higher for students from rural households, where the dropout risk rose by 1.6 percentage points, equivalent to 22 percent of the mean in municipalities less exposed to FARC-EP violence. By contrast, for urban households, conflict de-escalation has a non-statistically significant effect on the dropout risk.

To interpret these counter-intuitive findings, we examine the economic changes that occurred in the areas where the ceasefire occurred. In particular, the peace agreement brought about a notable increase in coca crops, which could have triggered the increase in dropout rates. This increase, according to Mejia, Prem, and Vargas (2019), was produced by “perverse incentives” related to the peace agreement, which established that monetary incentives would be granted to peasants that substitute coca for other legal crops. In fact, we find that most of the increase in dropout risk after the ceasefire is explained by the municipal potential of coca leaf production in the areas where the students’ households are located. Finally, using georeferenced information on rural schools and coca crops for the periods 2010–2012 (before the ceasefire) and 2013–2016 (after the ceasefire), we estimate the effect of the proximity in kilometers to coca crops on the dropout risk for both time periods in municipalities highly exposed to FARC-EP violence. We found that for the first period (before the ceasefire), proximity to coca crops was negatively correlated with dropout risk, whereas it was positively correlated with dropout risk for the second period (after the ceasefire).

Based on the existing literature on civil war and educational outcomes, it was expected that the de-escalation of the conflict with the FARC-EP would prompt improvements in the educational trajectories of individuals, particularly those from the municipalities most heavily affected by the armed confrontation. In fact, recent literature (Namen, Prem, and Vargas, 2021)[40] has examined how the recent peace process with the FARC-EP improved educational indicators at the municipality level (akin to a county) in Colombia. Nonetheless, the sizable increase in coca crops driven by the peace negotiations (Mejía, Prem and Vargas 2019) apparently provoked a result in the opposite direction. Consequently, children and youth from highly violent municipalities attracted by the high short-term returns would become laborers for coca crops and would drop out of school.

This research contributes to the existing literature in at least three ways. First, it provides evidence on the potential effects of the de-escalation of armed conflicts on the educational trajectories of individuals from the poorest households. Indeed, the literature that analyzes the effect of the decrease in violence in armed conflict on education in the short run is scarce. Most of the research focuses on estimating the conflict's long-term impact on different indicators of socioeconomic development

[40] Namen, Prem, and Vargas (2021), using a difference-in-differences methodology, evaluate the effect of the de-escalation of the Colombian conflict on school dropout rates and find a 19 percent reduction in dropout rates after the ceasefire in municipalities formerly affected by FARC violence. They show evidence that supports that the main mechanism is reduction in victimization after the cease-fire. The main difference with our paper is that our data allows us to track the educational trajectories of each child and teenager in the Colombian school system. Meanwhile, Vargas et al. (2020) uses the C-600, which enables them to see the total number of students enrolled in each grade at the beginning of the school year and the total number of students at the end. There are three potential limitations with the C-600 data: First, the calculus of a municipality-level weighted average of the school-specific dropout rates is giving greater importance to large schools over small ones, and the larger ones are mainly urban schools. Second, they see the change in the total number of students, but they cannot distinguish if a student is new, if he dropped out, if he changed schools, or if he migrated. Last, they capture the intra annual absenteeism, but they are not able to calculate the annual dropout rate.

(Chen, Loayza, and Reynal-Querol, 2007; Hoeffler and Reynal-Querol, 2003; Miguel and Roland, 2005; Stewart, Huang, and Wang, 2011). A paper closely related to the subject of this research is Chamarbagwala and Morán, (2011), who find evidence of a short-term decrease in school dropout rates after the end of the Guatemalan civil war.

Second, our research contributes to the understanding of the relationship between armed conflicts, illicit economic activity, and educational trajectories. The de-escalation of conflicts and even the disappearance of rebel forces who financed themselves through illicit activities do not necessarily imply a reduction in the role played by such activities in local economies and their effect on the inhabitants of these territories. The results obtained contribute to the literature that highlights the importance of local and regional contexts in educational trajectories (Chetty, Hendren, and Katz, 2015, García et al., 2015), as well as to the literature that evaluates how changes in labor markets affect the human capital accumulation of children and teenagers (Sviatschi, 2018; Shah and Steinberg, 2017; Cascio and Narayan, 2015). Notwithstanding, the negative relationship between school attainment and booms of primary products have been studied in depth. One example is the case of coffee booms in Colombia, analyzed by Carrillo (2020). To the best of our knowledge, our work is the first to examine this issue in the context of a de-escalation of violence.

Third, this paper contributes to recent academic work that studies the consequences of ending the five-decades-long Colombian domestic conflict. Some of these papers emphasize the unintended consequences and perils of the peace negotiations (2012–2016), including the rise in murders of local leaders in places previously dominated by the FARC-EP (Prem et al., 2019), the intensification of deforestation (Prem, Saavedra and Vargas, 2020), and the boom in coca crops (Mejía, Prem and Vargas and 2019).

The rest of this document is organized as follows: Section 4.2 introduces the context of the

Colombian armed conflict and the central elements regarding the peace negotiations with the FARC-EP. Section 4.3 describes the data used and presents descriptive statistics of the different variables of interest. Sections 4.4 to 4.7 present the identification strategy, the main results, and discusses the mechanisms. Finally, Section 4.8 presents the conclusions.

4.2 Context

Violence has been intermittently and asymmetrically present throughout Colombia's history. The Colombian domestic conflict has been particularly long and devastating. Social scientists locate its roots in the 1940s and 1950s (Fals-Borda, Guzmán-Campos, and Umaña-Luna, 2005; Pizarro, 2011; Sánchez and Meertens, 2011). Figures from Colombia's National Center for Historical Memory indicate that between 1985 and 2012,[41] the domestic conflict left six million forcibly displaced people and almost 220,000 killed, not to mention the victims of child recruitment, kidnappings, land dispossession, and sexual abuse, among other crimes.

Most authors trace the origins of Colombia's civil conflict back to *la Violencia*, a ten-year civil war that lasted from 1948 to 1958, involving mostly rural people who fought against each other under the banner of their political parties—the Liberal and the Conservative. After a decade of confrontations, violence decreased with the National Front, a political agreement between the two parties to alternate power. The loose ends of the National Front and the inefficiency of the government to foment rural development and to provide broad access to land for landless peasants created the conditions for the emergence of several guerrilla groups, including the FARC-EP, the National Liberal Army (ELN), the Popular Liberation Army (EPL), and the April 19 Movement (M-19) (Villamizar, 2017).

[41] This period of time would be the last 27 years of a nearly 50-year conflict.

In 1966, two years after a large military operation against armed liberal peasants known as the Attack on Marquetalia, the FARC-EP was officially founded as an offensive peasant guerrilla army. Likewise, in the mid-1960s, the ELN was formed by liberal guerrilla fighters who were involved in *la Violencia*, along with students and professors inspired by the Cuban Revolution (CNMH, 2013). During the 1960s and 1970s, the intensity of the armed conflict remained low, but by the 1990s, peripheral armed confrontations devolved into an intense war (Meernik, Demeritt, and Uribe-López, 2019).

In 1990, the expansion of the guerrilla forces and their financing strategies had a direct impact on daily lives of the regional and local elites (Arjona, 2016). In response, these elites formed paramilitary groups, which would later be backed by drug-trafficking groups (Duncan, 2013). Rebel groups got involved in illicit activities to finance their armed activities (Kalyvas, 2015), including taxing and planting of coca crops (Sánchez and Díaz, 2007). Empirical evidence has shown that Colombia's coca crops have served as funding sources for the armed groups and have reinforced the existing conflict (Angrist and Kugler, 2008; Mejía and Restrepo 2013; Ross, 2004).

Living in conflict-affected areas is a life-changing experience that modifies household's decisions and has lasting consequences.[42] Among other effects, exposure to conflict increases the probability that school-age children will drop out of school and enter the labor force prematurely (Rodriguez and Sánchez, 2012). It makes it more difficult for households to generate income and maintain consumption levels (Ibáñez and Moya, 2010). It also has psychological consequences that can lead to endemic poverty (Moya, 2018) and changes the choice of economic activities to less profitable ones (Arias, Ibáñez, and Zambrano, 2019).

[42] One of five young residents in areas most affected by armed conflict in Colombia have been victims of kidnapping, forced internal displacement, extortion, murder of a close relative or threat (Bedoya et al., 2019).

In 2011, after more than 50 years of armed confrontation, Colombian Government and the largest guerrilla FARC-EP began a secret exploratory phase of peace talks (Villamizar, 2017). The peace process began officially in 2012, and it took four years to reach an agreement. The negotiating agenda included six main issues: rural reform, political participation, termination of conflict (ceasefire, disarmament, demobilization, and reintegration), strategy against illicit drugs, victims' reparations, and implementation and verification mechanisms (Meernik, Demeritt, and Uribe-López, 2019). Thus, the Colombian case provides a unique setup that enables the effects of the de-escalation of military actions on school attendance to be determined. In particular, two critical moments in the peace process changed the dynamics of the conflict. First, in December 2012 the FARC-EP declared a unilateral ceasefire as a credible sign of commitment. According to data provided by Colombia's National Center of Historical Memory (CNMH), during 2013 and until the signature of the peace agreement in 2016, the number of victims associated with FARC-EP violence decreased by 80 percent. Second, in May 2014 the delegates from both the national government and FARC-EP established the parameters for the substitution of illicit crops and for a program to promote alternative agrarian development. The program sought to decrease coca crops by means of voluntary substitution in exchange for cash transfers, technical assistance, and grants for productive projects (Colombian National Government, 2016). The agreement on the issue was announced in the national press and generated anticipatory effects. Unexpectedly, the prospect of the future rents derived from coca substitution created perverse incentives and caused an unprecedented surge in illicit coca production. Thus, between 2013 and 2015, the number of hectares planted with coca increased from 47,788 to 146,140 (Mejia, Prem, and Vargas, 2019[43]; Saavedra Ladino and Wiesner, 2019).

[43] Mejía, Prem, and Vargas (2019) exploit a unique historical setting to study the anticipatory effect, generated by the announcement of a future policy, on the production of cocaine in Colombia. After two years of formal negotiations between the Colombian government and the largest of its opposing guerrilla forces, FARC-EP, a combined effort on the part of all parties to the negotiations created a strategy against illicit coca crops. In 2014 they announced that after the peace agreement, the government would provide monetary incentives for coca-growing communities to voluntary substitute coca for legal crops. To causally identify the effect of the announcement in coca-suitable municipalities of the

4.3 Data

Throughout the empirical analysis carried out in this work, we use several administrative records and municipal-level data that come from different sources and can be grouped as follows:

*Data on school dropout rates***.** To examine the educational trajectories of children and youth in school, we use the anonymized annual census of primary and secondary students in Colombia (SIMAT). This core database of the Ministry of Education covers the universe of students in Colombian schools during the period 2008–2016. Each year of this dataset compiles information on all students enrolled in Colombian schools at an individual level. Therefore, we can track the educational trajectories of each child and youth in the Colombian school system. To analyze school dropout risk, we construct a discrete variable, which takes the value of zero during the years the child or youth is attending school and one if the child or youth stops attending school before 11th grade – the last grade of secondary education (see Table C.1 in Appendix for the structure of the educational data). These variable measures the likelihood of experiencing a dropout event, which henceforth will be referred to as the dropout risk.

Of the approximately 1,400,000 students tracked during the years prior to the ceasefire (2009–2013), 6.06 percent dropped out of school each year and about 30 percent attended a school located in municipalities heavily affected by FARC-EP violence. Furthermore, in line with the empirical evidence, the yearly dropout risk was higher for students living in municipalities strongly affected by FARC-EP violence (7.25 percent) than for students in less affected municipalities (5.58 percent).

---

expansion of coca crops, the authors implemented a difference-in-differences methodology with municipality fixed effects, department-year fixed effects, and a large set of controls. They find that the announcement of a future policy caused an unprecedented expansion of coca crops. Likewise, if the estimated probability that a municipality will receive the substitution incentive scheme increased by 50 percent relative to the mean (from 5 to 7.5 percent), then the incidence of coca crops would increase by 0.57 standard deviations.

*Household Data***.** To identify siblings belonging to the country's poorest households, we merge the anonymized administrative information of students with a Colombian means test database named SISBEN (Identification System for Subsidies Beneficiaries). SISBEN was created in 1994 as a targeting tool for social programs. In general, municipal officials conduct door-to-door interviews in neighborhoods of strata below level four[44] (Camacho and Conover, 2011) collecting information on demographics, housing characteristics, income, and employment at the individual and the household level. This information is processed to calculate and assign a poverty index or score based on long-term living conditions. SISBEN allows us to identify siblings since it provides information of all members of a given household.

We use the second SISBEN wave collected from 2005 to 2008. This database reported rich demographic and sociodemographic information of almost 32 million individuals, the poorest Colombians, and comprised around 75 percent of the total population. The merge of SISBEN with SIMAT data enables a comparison of school attendance across siblings of different ages.[45]

*Conflict Data***.** To trace the patterns of conflict among the municipalities exposed to FARC-EP violence, we use a municipal-level conflict dataset compiled by the CNMH. The CNMH was created by the Law 1448/2011 with the mission of reconstructing the history of the Colombian internal conflict and to fulfill for the victims of the conflict their right to know the truth. Its main objective is to collect and organize the narratives and events that affected the victims of the conflict. In particular, the CNMH's dataset compiles the microdata of violent events[46] occurred since the mid-1950s, registering the municipality where it took place, the year of occurrence, the number of victims

[44] Colombia neighbourhoods are geographically stratified into six levels, in particular stratum 1 the poorest and 6 the wealthiest. (Camacho and Conover, 2011).

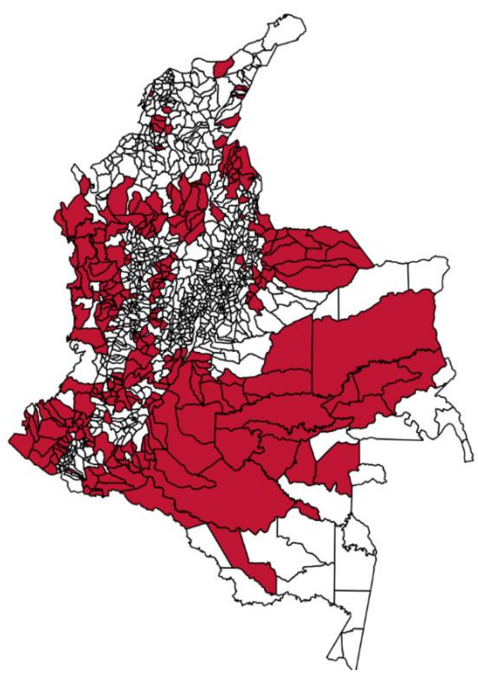
[45] We identify siblings as those who live under the same household and report having the same household head as father or mother.

[46] Has recorded information on Colombia's conflict since 1958.

involved, the type of violent event, and the identity of the perpetrator.

Based on this information, we calculate the total number of victims[47] and the victims' rate (per 100,000 people) affected by the FARC-EP's violent actions. Then, we categorize the municipalities highly exposed to FARC-EP violence as those where its victim rate per 100,000 inhabitants was higher than the national average prior to 2013 (Figure 4.1).

Figure 4.1: Municipalities with High Exposure to FARC-EP Violence



Notes: The figure shows the municipalities highly exposed to FARC-EP violence. Especially, those where the FARC-EP victims' rate per 100,000 inhabitants is higher than the national average prior to 2013. Source: CNMH.

*Area with coca crops***.** We compile the number of hectares of illicit coca crops from the database of municipal variables of Economic Development Research Center (CEDE) at the University of the

[47] The victims are defined as those who suffered one or many of the following actions (in their own self or in relation to a family member): selective killings, forced disappearances, attacks on civilian population, terrorist attacks, sexual violations, recruitment, massacres, war actions and kidnappings.

Andes. The CEDE database contains municipal information of illicit coca crops from the Illicit Crops Monitoring Program - SIMCI- and Colombia's Observatory on Drugs. This dataset registers the number of hectares of coca cultivated in each municipality. Based on this information, we construct a dummy variable equal to one if the municipality had coca crops prior to the 2013 ceasefire.

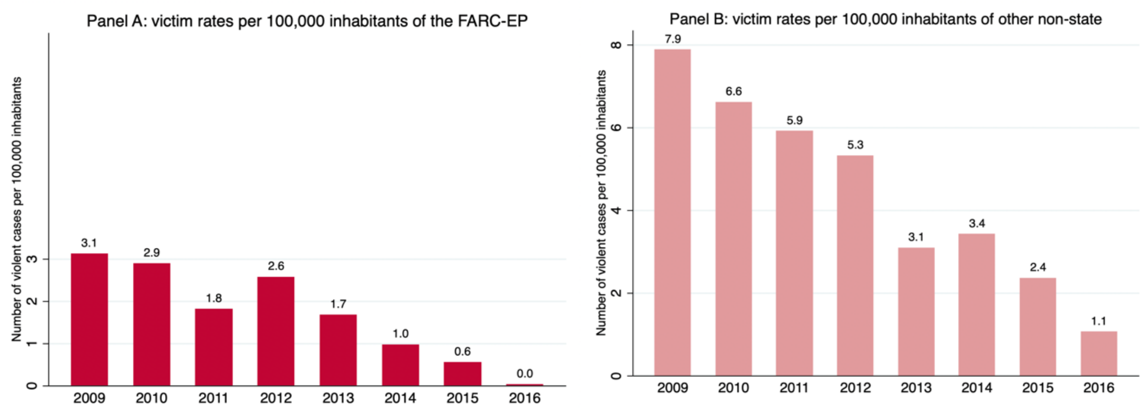
To rule out the likely endogeneity of the number of hectares of coca crops, we calculate the cross-municipal variation of the potential production of coca. First, we use a coca suitability index estimated by Mejia and Restrepo (2015) based on information from a nationally representative household survey of coca farmers. This survey provides self-reported data on coca crops, which they merge with exogenous municipal geographic and weather characteristics. Second, we obtain the potential municipal area suitable for coca crops as the product of the coca suitability index–standardized between zero and one–and the municipality's area. Third, we create a dummy variable equal to one if the municipal potential suitable area for coca cultivation is higher than the national median.

Finally, since municipalities are the lowest level of aggregation in the reports of the number of coca hectares cultivated, we also use coca density maps for the period 2009–2016 from the SIMCI. In 1999 the United Nation Office of Drugs and Crime (UNODC) and the Ministry of Interior and Justice of the Government of Colombia started SIMCI with the purpose of annually processing satellite images to update statistics on the location and extension of coca fields. This data allows us to compare the impact of the proximity to coca crops.

Descriptive statistics

Figure 4.2 presents the victim rate per 100 thousand inhabitants of the Colombian conflict. As observed, the conflict de-escalated after the Central High Command of the FARC-EP ordered a unilateral ceasefire in December 2012 for 60 days (El Tiempo, 2015). As a result of the ceasefire the FARC-EP violent action rate dropped sharply from 2.6 in 2012 to 0.0 in 2016 –a decrease by 100 percent. Furthermore, the violent action rate of all non-state armed groups fell by 79 percent.

Figure 4.2: Victims' Rate of Conflict Actions at National Level

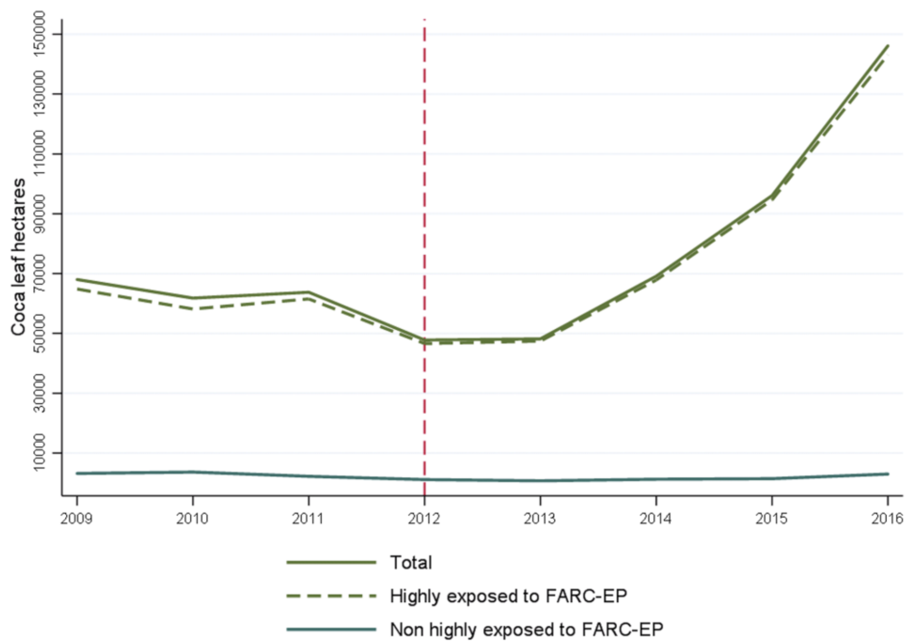


Notes: The figure shows the victim rates per 100,000 inhabitants of the FARC-EP in panel A and in panel B other non-state armed groups, as paramilitary, other guerrilla group -ELN-, some dissidents of armed groups and organized crime groups, victim rates. Violent events producing victims include selective murder, attacks on population, terrorist attacks, forced disappearance, kidnapping, massacres, sexual violence, recruitment, and anti-personnel mines. Source: National Center of Historical Memory.

In 2014, during the peace negotiations, the government announced the coca voluntary substitution program. The program would grant monetary transfers to households and peasants that undertake substitution of coca crops. Thus, for signing the agreement a peasant would receive COL$ 2.0 million (2.5 monthly minimum wages), COL$ 2.0 million every two months after eradication, COL$ 20.0 million for the new productive project and COL$ 3.2 million for technical assistance (Junguito-Bonnet, Perfetti del Corral, and Delgado-Barrera 2017; ONUDC, 2018). The announcement of the policy generated an unintended anticipatory effect increasing coca cultivation (Mejía, Prem, and

Vargas 2019; Saavedra Ladino and Wiesner, 2019). Figure 4.3 presents the evolution of coca crops from 2009 to 2016 for all the country and for municipalities highly and moderately exposed to FARC-EP violence, showing how coca crops expanded. From 2009 to 2013, coca cultivation declined and reached an all-time low (UNODC, 2013). Nevertheless, in just three years, by 2016 the land cultivated with coca exceeded the historical levels. Moreover, it is important to underscore that nearly 96 percent of coca plantations are in municipalities highly exposed to FARC-EP violence.

Figure 4.3: Coca Hectares Cultivated Nationwide (2009-2016)



Notes: The figure shows the hectares of coca cultivated in Colombia disaggregated by the total national and the total plantations located in municipalities highly and not highly exposed to FARC-EP violence. Source: records from SIMCI and Colombia`s Drogue Observatory complied by the University of the Andes in the Municipal CEDE panel.

Table 4.1 Panel A shows summary statistics for all children enrolled in the school system between 2009 and 2016 whose siblings were also attending an educational institution in this period and who belong to the country's poorest households. Regarding exposure to FARC-EP violence, students who attended school in municipalities with higher exposure to violence had a higher risk of dropping out (7 percent before and 9 percent after ceasefire) compared to those with lower exposure (5 percent before and 7 percent after ceasefire), before and after the unilateral ceasefire. Likewise, we find that rural children and youth were more likely to abandon school, and that the school dropout gap between rural and urban students was higher for students exposed to FARC-EP violence.

Overall, the ceasefire occurred during a period of decline in school attendance in Colombia. Nevertheless, school dropout trends suggest that more students in higher grades are expected to leave school regardless of the level of violence. Therefore, the rise in the school dropout could reflect this trend. Finally, Panel B focuses the analysis at the municipal level. It provides evidence that the illicit number of hectares planted with coca crops was higher in municipalities highly exposed to FARC-EP attacks (Angrist and Kruger, 2006). It is important to highlight that school children in municipalities highly exposed to FARC-EP violence were exposed to less violent attacks after ceasefire (21 victims per 100,000 inhabitants, column (6) vs. 6 victims per 100,000 inhabitants, column (8). However, they also are exposed to more hectares of illicit cultivation (155 hectares of coca, column (6) vs. 237 hectares of coca, column (8) after the first FARC-EP unilateral ceasefire).

Table 4.1: Descriptive Statistics of Students and Municipalities

	Panel A: students' characteristics							
	Municipalities **not** highly exposed to FARC-EP violence				Municipalities highly exposed to FARC-EP violence			
	Before the unilateral ceasefire (2010–2012)		**After the unilateral ceasefire (2013–2016)**		**Before the unilateral ceasefire (2010–2012)**		**After the unilateral ceasefire (2013–2016)**	
	Obs. [1]	Mean/ (Std. Dev.) [2]	Obs. [3]	Mean/ (Std. Dev.) [4]	Obs. [5]	Mean/ (Std. Dev.) [6]	Obs. [7]	Mean/ (Std. Dev.) [8]
Annual dropout	3,1078,35	0.055	2,272,198	0.074	1,213,086	0.072	940,529	0.087
		(0.00013)		(0.00017)		(0.00023)		(0.00029)
Annual dropout urban	2,109,435	0.053	1,504,319	0.074	671,517	0.066	512,971	0.0802
		(0.00015)		(0.00021)		(0.00030)		(0.00037)
Annual dropout rural	998,400	0.060	767,879	0.075	541,569	0.080	427,558	0.096
		(0.0002)		(0.00030)		(0.00036)		(0.00045)
Grade	3,107,835	6.19	2,272,198	6.95	1,213,086	5.71	940,529	6.471
		(0.0016)		(0.0018)		(0.0027)		(0.00287)
Sex (female=1)	3,107,835	0.51	2,272,198	0.5	1,213,086	0.50	940,529	0.49
		(0.00028)		(0.0003)		(0.0004)		(0.0005)
Age	3,107,835	12.53	2,272,198	13.55	1,213,086	12.30	940,529	13.25
		(0.0018)		(0.0019)		(0.003)		(0.0032)

	Panel B: Characteristics of the municipalities							
	Municipalities **not** highly exposed to FARC-EP violence				Municipalities highly exposed to FARC-EP violence			
	Before the unilateral ceasefire (2010–2012)		**After the unilateral ceasefire (2013–2016)**		**Before the unilateral ceasefire (2010–2012)**		**After the unilateral ceasefire (2013–2016)**	
	Obs. [1]	Mean/ (Std. Dev.) [2]	Obs. [3]	Mean/ (Std. Dev.) [4]	Obs. [5]	Mean/ (Std. Dev.) [6]	Obs. [7]	Mean/ (Std. Dev.) [8]

Victims attributed to FARC-EP over 100.000 inhabitants	2,368	0.27	2,603	0.20	984	20.93	1,022	5.53
		(0.050)		(0.060)		1.19		(0.54)
Hectares of coca	3,004	3.40	3,004	2.15	1,484	155.76	1,484	237.92
		(0.52)		(0.51)		11.97		(26.74)

Notes: The table shows in Panel A the average age, grade, percentage of girls, and the general, urban, and rural percentage of school dropouts for municipalities not highly and highly exposed to FARC-EP violence, before and after the ceasefire. Also, in Panel B it presents the municipal difference in the average of victims and coca-growing hectares for municipalities not highly and highly exposed to FARC-EP violence, before and after the unilateral FARC-EP ceasefire. Sources: SIMAT, CEDE, CNMH, and SISBEN.

4.4 The Effect of Conflict De-escalation on School Dropout Rate

This section examines the effect of the unilateral FARC-EP ceasefire on education. First, we describe the identification strategy, and then we analyze the relationship between the de-escalation of Colombia's conflict and the school dropout risk, as well as the mechanisms underlying this relationship.

Baseline Econometric Specification

To determine the consequences of the Colombian peace process on households' educational decisions in areas of high violent presence of FARC-EP, we employ a *difference-in-differences* methodology with household and year fixed effects, using information on children and youth from 487,485 non-migrants' households[48] between 2009 and 2016.

Using sibling microdata, our main specification exploits the unexpected timing of the first unilateral FARC-EP ceasefire, announced on December 2012, as well as the geographic variation of the exposure to FARC-EP violence across municipalities prior to this event. Although the educational trajectories and characteristics of students attending school in territories highly affected by FARC-EP violence differ from the trajectories of those who study in less violent territories, these differences are considered by households fixed effects. Thus, by including household fixed effects, we control for siblings' common family backgrounds and parents' preferences.

We estimate the following difference-in-differences model:

$$y_{imht} = \beta_0 + \beta_1 CEASE_t * FARC_m + \alpha X_{it} + \delta_h + \mu_t + \varepsilon_{imht} \quad [4.1]$$

where y_{imht} is a dummy that takes the value of one if student *i* from household *h*, dropped out of school in year *t*, given he or she was attending school in the previous year *t-1*. $CEASE_t$ is a dummy

[48] We restrict our sample to non-migrants' households to solve potential biases coming from differential trajectories that displaced households may have (Becker, et al., 2020; Oyelere and Wharton, 2013; Ibáñez and Velásquez 2008)

that takes the value one since 2013 after the first unilateral FARC-EP ceasefire took place. $FARC_m$ is a dummy that takes the value of one for those highly exposed to FARC-EP violence municipalities prior to the ceasefire. The interaction between $CEASE_t$ and $FARC_m$ identifies the municipalities experiencing ceasefire, where the number of victims associated with FARC-EP violent actions reduced during 2013–2016. X_{it} are student characteristics such as age, sex and grade. δ_h are households fixed effects and μ_t are time fixed effects. Finally, ε_{ihmt} is the error term clustered at the household level.

The estimates of this strategy determine the effects of ceasefire on dropout decisions of children and youth from municipalities with high exposure to FARC-EP violence in years following the ceasefire relative to previous years. In particular, the coefficient β_1 is the difference-in-differences estimate of the impact of de-escalation of the conflict during the years that the FARC-EP ceased their attacks, controlling for the differential effect of conflict de-escalation across siblings in the same household. Interpreting β_1 as the causal effect of conflict de-escalation does not require one to assume that the ceasefire timing within households was random. Instead, it involves the weaker assumption that, conditional on the baseline controls, the de-escalation of the conflict is not correlated with unobserved time-varying household characteristics that affect decisions about education. Before presenting the main results, the next section presents evidence to reject the hypothesis of any trend differences between the treatment and the control groups before 2013.

Identification Assumption

The main identifying assumption of our baseline specification is that there exists a common trend across municipalities experiencing different levels of FARC-EP violence in the absence of the ceasefire. This assumption could be violated if, for instance, the risk of dropping out of school was increasing in municipalities highly affected by FARC-EP violence before 2013. We address this concern by including the prior year (2010–2012) in the main specification, as shown in the following equation:

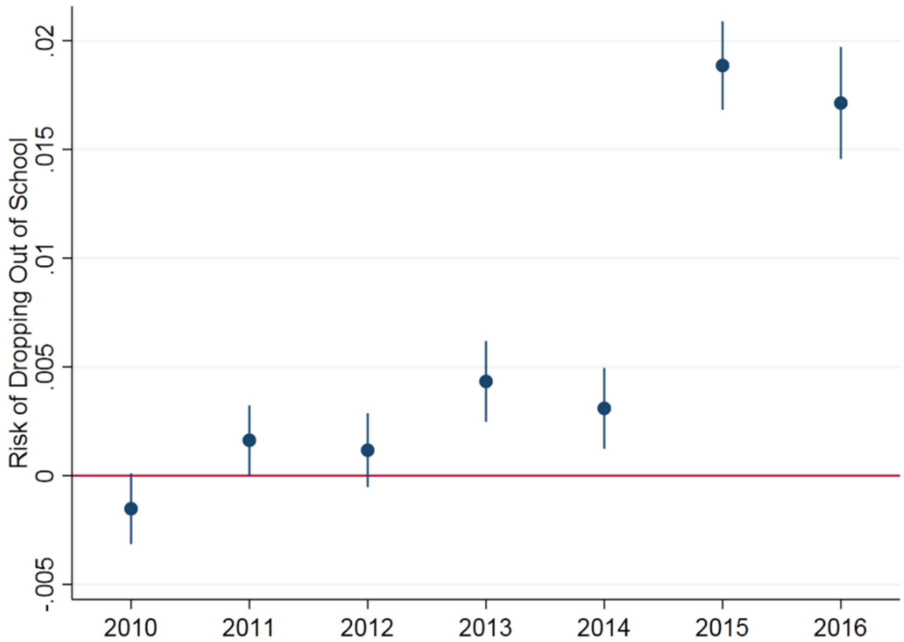
$$y_{ihmt} = \beta_0 + \sum_{t=2010}^{2016} \beta_t PRE - CEASE_t * FARC_m + \sum_{t=2013}^{2016} \beta_t CEASE_t * FARC_m$$

$$+\alpha X_{it} + \delta_h + \mu_t + \varepsilon_{ihmt} \qquad [4.2]$$

$PRE - CEASE_t * FARC_m$ stands for the interaction of municipalities with high exposure to FARC-EP violence and time dummies for all periods before the de-escalation of the armed conflict. Likewise, $CEASE_t * FARC_m$ is where the areas highly exposed to FARC-EP violence interacted with time dummies for all periods after the ceasefire. The parameters β_t can be interpreted as the gap in the risk of dropping out school between municipalities highly and not highly exposed to FARC-EP violence, in the years preceding the decline of violence relative to the year 2009.

Figure 4.4 reports the coefficients resulting from estimating equation [4.2]. First, it is evident that before the ceasefire the coefficients are not statistically significant. This result confirms the absence of differential dropout risk between students from municipalities highly exposed to FARC-EP attacks and those that were not highly exposed relative to the year 2009. Second, the magnitude and significance of the estimators increase after the large surge of coca crops in 2014.

Figure 4.4: Effects of Conflict De-escalation on School Dropout by Year



Notes: This figure presents the β_t coefficient from equation [4.2]. Thus, the risk of dropping out of school increased for children and youth who attended school in municipalities highly affected by FARC-EP violence.

4.5 Results

In Table 4.2 we report the coefficients of main specification in equation [4.1]. Column (1) presents the results of the baseline specification, where we identify municipalities highly exposed to FARC-EP violence as those where the victims' rate per 100,000 inhabitants was higher than the national average during 2008–2012. The estimates suggest that the de-escalation of the Colombian conflict led to a 0.9 percentage point increase in the dropout risk for students who attended schools in municipalities highly exposed to FARC-EP violence. This translates into a 13 percent increase relative to the mean of that variable in municipalities not highly exposed to FARC-EP violence after the ceasefire. These results are robust to alternative definitions of "municipalities highly exposed to FARC-EP violence" (see Table C.1 in the Appendix).

Thus, the FARC-EP's ceasefire had negative effects on school attendance of the most vulnerable children and teenagers in highly violent municipalities. Subsequently, civilians living in conflict zones face a puzzling reality, especially when illicit activities are present. Between 2013 and 2016, the area planted with coca crops doubled, and the number of victims associated with the FARC-EP fell by 100 percent, likely lowering the cost of, and increasing the returns to, working on coca crops. Our results are consistent with empirical evidence from Peru suggesting that an increase in coca leaf production in districts that had optimal conditions to grow this plant has large effects on children's trajectories of human development since it decreases the number of years they remain in school (Sviatschi, 2018). It is likely that the rise in school dropout risk found is driven by the inclusion of students attending school in large cities with low exposure to FARC-EP violence. Thus, we estimate equation [4.1] excluding households from the 13 largest Colombian cities[49] and also restricting the sample to municipalities with a population of less than 500,000 inhabitants in 2012. We find that the coefficient of the de-escalation of conflict is larger in both cases (See columns 2 and 3).

[49]The 13 largest Colombian cities are: Bogotá, Cali, Cartagena, Pasto, Manizales, Barranquilla, Montería, Bucaramanga, Cúcuta, Medellín, Villavicencio, Pereira, and Ibagué.

Table 4.2: Effect of Conflict De-escalation on School Dropout Risk

Dependent variable: Risk of dropping out of school

	Highly exposed measure by mean				
	General [1]	Without big cities [2]	Restricted to municipalities with fewer than 500,000 inhabitants in 2012 [3]	Students from urban households [4]	Students from rural households [5]
CEASE x FARC	0.00927***	0.0115***	0.0118***	0.000919	0.0167***
	(0.000584)	(0.000606)	(0.000862)	(0.000717)	(0.000982)
Observations	5,510,345	4,394,027	2,120,151	3,666,590	1,843,755
Students	1063622	842563	407485	712766	350856
Households	487485	381581	181401	333707	153778
R-squared	0.216	0.215	0.216	0.218	0.217
Households FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0708	.066	.0665	.0708	.0707

Notes: This table presents the results from the main specification in equation [4.1]. *Highly exposed measure by mean* is a discrete variable equal to one for municipalities where the mean of the FARC-EP victim rate per 100,000 inhabitants between 2008–2012 was higher than the national average. *CEASE* is a dummy variable equal to one for the period since 2013. All regressions include sex, grade (in log), age, and age squared as controls. Robust standard errors area clustered at the household level. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. P-values for standard errors in parenthesis: * significant at 10%, ** significant at 5% and *** significant at 1%.

Next, we estimate the effects of the decline in FARC-EP violence on urban and rural children and youth. For students from rural households heavily affect by conflict, the risk of dropping out of school increases by 1.6 (column 5) percentage points. This translates into a 24 percent increase relative to the mean of municipalities not highly exposed to violence after the ceasefire. Additionally, we document that in urban areas, conflict de-escalation does not affect households' schooling decisions. Therefore, the findings clearly show that changes in conflict dynamics are more likely to affect rural children and youth.

Finally, in Table 4.3 we report the coefficients of our main specification equation [4.1] by level of education—primary and secondary— students are attending. For students from municipalities highly exposed to FARC-EP violence in primary school, the risk of dropping out school increased by 0.3 percentage points, or 5.5 percent from the mean of the control municipalities after 2012 (column 1). For students in secondary school from the same municipalities, the risk of dropping out increased by 1.2 percentage points, or 15 percent from the mean of the control municipalities after the ceasefire (column 2). As children should attend primary school between 6 and 10 years of age and secondary school between 11 and 16 years of age, the increase in the risk of dropping out was concentrated among the older students. Similar results are reported by Sviatchsi (2018), who shows that the expansion of the cocaine industry in Peru caused secondary school students to drop out.

In column (3) of Table 4,3, we report the effect of conflict de-escalation on dropout risk for urban students in primary education. For urban primary school students, the risk of dropout decreased by 0.6 (column 3) percentage points, equivalent to 11 percent from the mean of the control places after the ceasefire. In contrast, for rural primary school students from municipalities highly exposed to FARC-EP violence, the risk of dropout increased by 1.2 percentage points, or 24 percent from the mean of the municipalities not highly exposed to FARC violence (column 5). Likewise, for rural students in secondary education from the same municipalities, the risk of dropout increased by 25 percent from the mean of the municipalities that were not exposed to FARC-EP attacks after 2012 (column 6).

Table 4.3: Effect of Conflict de-escalation On School Attainment by Level of Education

Dependent variable: Risk of dropping out of school

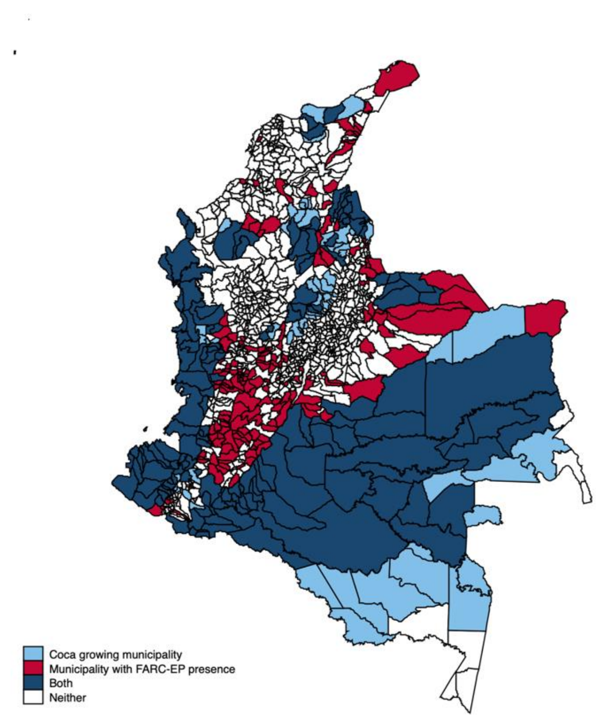
Highly exposed measure by mean						
	General		Students from urban households		Students from rural households	
	Primary [1]	Secondary [2]	Primary [3]	Secondary [4]	Primary [5]	Secondary [6]
CEASE x FARC	0.00330***	0.0120***	-0.00607***	0.00375***	0.0128***	0.0212***
	(0.000921)	(0.000769)	(0.00121)	(0.000911)	(0.00140)	(0.00140)
Observations	1,951,185	3,501,130	1,182,171	2,443,549	769,014	1,057,581
Students	512559	897146	316894	624064	195665	273082
Households	314551	452860	204512	315946	110039	136914
R-squared	0.306	0.335	0.311	0.335	0.303	0.337
Households FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0545	.0777	.0567	.076	.0508	.0822

Notes: This table presents the results from the main specification in equation [4.1]. *Highly exposed measure by mean* is a discrete variable equal to one for municipalities where the mean of the FARC-EP victim rate per 100,000 inhabitants between 2008–2012 was higher than the national average. *CEASE* is a dummy variable equal to one for the period after 2013. All regression includes sex, grade (in log), age, and age squared as controls. Robust standard errors area clustered at the household level. The mean of the controls is the mean of the not highly exposed to FARC violence after 2012. P-values for standard errors in parenthesis: * significant at 10%, ** significant at 5% and *** significant at 1%.

4.6 Mechanisms: Conflict De-escalation and Coca Crops

The results obtained reveal that the de-escalation of the armed conflict in Colombia has negatively impacted school attendance of students living in the municipalities historically affected by the violent actions of the FARC-EP. To understand such puzzling results, we analyze the surge of coca crops in the municipalities with a previous high incidence of FARC-EP where violent events fell sharply at the beginning the ceasefire. Figure 4.5 shows that 72 percent of coca-growing municipalities experienced previous high exposure to FARC-EP violence. Using this information, we test whether the effect of conflict de-escalation is more prevalent in municipalities highly suited to coca crop production.

Figure 4.5: Municipalities Exposed to FARC-EP Violence and Coca Crops



Notes: This figure presents a map of Colombia's municipalities. The dark red the areas define the territories highly exposed to FARC-EP violence where the FARC-EP victim rate per 100,000 inhabitants is higher than the national average. The light blue areas represent the coca growing municipalities and the dark blue areas are the municipalities with presence of both coca crops and previously high FARC-EP violence. Source: CNMH and CEDE's Municipal Panel 1993–2017.

The literature suggests that labor opportunities stemming from illicit economic activities—working in coca crops—increase the opportunity cost of attending school, which increases dropout risk (Angrist and Krugler, 2008; Bandera, Dehejia and Lavie-Rouse, 2015; Dammert, 2008). Moreover, the reduction in violence may also reduce the potential dangers of being involved in coca crop activities, which lowers the cost of working in this activity and incentivizes children and youth to take advantage of short-term high labor returns and to leave school. Consequently, a surge in coca crops increases the demand for low-skilled workers such as children and youth. Thus, when coca crops expand, child labor rises, and the relatively high earnings for participation in this labor market affects educational outcomes of students in areas suitable for coca growing (Sviatschi, 2018). Therefore, households' educational decisions in areas that were previously highly exposed to FARC-EP violence would in fact depend on the economic activities that flourished during the conflict period.

Qualitative evidence indicates that coca-related activities tend to affect children's educational decisions. For example, *El Espectador*, a national Colombian newspaper, reported in that in coca-growing areas, children leave school to harvest coca leaves in their family or neighbors' plantations. Additionally, in response to a question about how many families depended on coca, peasant in Tumaco (the municipality with the largest area of coca crops) replied "I would say every family" (*El Espectador*, 10/14/2019). Likewise, a young boy described to *El Espectador* his experience working in illicit crops, saying "once you get used to the routine, it is difficult to turn back, since you realize that coca is more profitable than studying" (*El Espectador,* 05/07/2019).

To test whether the surge in coca crops is the driving factor behind the increase in dropout rates in municipalities that experienced a reduction in FARC-EP violence, we introduce into our baseline specification a variable that indicates whether the municipality where the student is attending school is suitable for growing coca. Therefore, we estimate the following model:

$$y_{ihmt} = \beta_0 + \beta_1 CEASE_t * FARC_m + \beta_2 CEASE_t * FARC_m * COCA_m + \alpha X_{it} + \delta_h + \mu_t + \varepsilon_{ihmt} \quad [4.3]$$

Building from our baseline model we add the variable $COCA_m$, which is a dummy indicating whether the municipality's potential area to grow coca is higher than the country's median. In this case, the interaction between $CEASE_t$, $FARC_m$ and $COCA_m$ identifies the territories with both the highest reduction of FARC-EP violence during 2013–2016 and the largest areas suitable for coca growing.

Table 4.4 shows that the increase in the students' risk of dropping out of school is explained to a large extent by the interaction $CEASE_t * FARC_m * COCA_m$. Thus, it seems that the change of child labor in coca crops is the driving factor behind the increase in school dropout. Table 4.4 show that after including the triple interaction $CEASE_t * FARC_m * COCA_m$ the coefficient of $CEASE_t * FARC_m$ becomes smaller than the coefficient reported in Table 4.2[50]. Hence, coca crop suitability explains 68 percent of the effect previously reported in Table 4.2. Additionally, as shown in Table 4.4 columns (2) and (3), if we exclude from the sample the 13 largest cities or restrict the sample to municipalities with a population less than 500,000 inhabitants in 2012, we obtain similar results. In columns (4) and (5), the sample is divided between urban and rural students. For rural students, the surge of coca crops entirely explains the effect of conflict de-escalation on school dropout. Similarly, the results suggest that for rural students, short-term returns to working in coca crops strongly incentivize school dropout, to a greater extent than for urban students.

[50] Results maintain if we measure *Highly exposed to violence municipalities by median:* a discrete variable equal to one for municipalities where the median of the FARC-EP victim rate per 100,000 inhabitants between 2008–2012 was higher than the national median. ceasefire -See table C.3 in appendix-

Table 4.4: Effect of Conflict De-escalation and Coca Surge on School Dropout Risk

Dependent variable: Dropping out of school

	General	without big cities	Restricted to the municipalities with a population in 2012 less than 500,000 inhabitants	Students from urban households	Students from rural households
Highly exposed measure by mean					
	[1]	[2]	[3]	[4]	[5]
CEASE x FARC	0.00348***	0.00357***	0.00421***	0.00432***	0.000904
	(0.000985)	(0.00103)	(0.00151)	(0.00119)	(0.00175)
CEASE x FARC x COCA	0.00678***	0.0116***	0.00743***	-0.00580***	0.0180***
	(0.00125)	(0.00128)	(0.00192)	(0.00152)	(0.00217)
Observations	5,510,345	4,394,027	2,120,151	3,666,590	1,843,755
Students	1063622	842563	407485	712766	350856
Households	487485	381581	181401	333707	153778
R-squared	0.216	0.215	0.216	0.218	0.218
Households FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Mean controls	.0708	.066	.0665	.0708	.0707

Notes: This table presents the estimation results of equation [4.3]. *Highly exposed measure by mean* is a discrete variable equal to one if the FARC-EP victim rate per 100, 000 inhabitants during 2008-2012 is higher than the national average. *CEASE* is a dummy variable equals one for the period after 2012. *COCA* is a dummy variable equal to one for municipalities in which their area suitable for coca cultivation is larger than the national average. All regressions include the following controls: sex, grade (in log), age, and age squared. Robust standard errors area clustered at the household level. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. P-values for standard errors in parenthesis: * significant at 10%, **significant at 5% and ***significant at 1%.

Finally, Table 4.5 reports the results of estimating equation [4.3] for primary and secondary education. For primary school students, the reduction of FARC-EP violence during 2013–2016 marginally lowered dropout risk by 0.2 percentage points; nonetheless, it increased due entirely to the surge of coca crops. For secondary school students, both coefficients—the reduction of FARC-EP violence and the surge of coca crops—are positively related to dropout. However, the latter explains 84 percent of the effect of conflict de-escalation on school dropout.[51] Moreover, the coefficient of interaction $CEASE_t * FARC_m * COCA_m$ is only statistically significant for rural students both at the primary and the secondary level.

Table 4.5: Effect of Conflict De-Escalation and Coca Surge on School Dropout Risk by Level of Education

Highly exposed measure by mean						
	General		Students from urban households		Students from rural households	
	Primary [1]	Secondary [2]	Primary [3]	Secondary [4]	Primary [5]	Secondary [6]
CEASE x FARC	-0.00307*	0.00261**	-0.00511**	0.00345**	0.00104	-0.000647
	(0.00160)	(0.00126)	(0.00205)	(0.00149)	(0.00255)	(0.00236)
CEASE x FARC x COCA	0.0114***	0.0111***	0.00101	-0.000408	0.0177***	0.0220***
	(0.00199)	(0.00162)	(0.00257)	(0.00192)	(0.00312)	(0.00302)
Observations	1,951,185	3,501,130	1,182,171	2,443,549	769,014	1,057,581
Students	512559	897146	316894	624064	195665	273082
Households	314551	452860	204512	315946	110039	136914
R-squared	0.306	0.335	0.311	0.335	0.304	0.337
Households FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0545	.0777	.0567	.076	.0508	.0822

Notes: This table presents the results of equation [4.3]. *CEASE* is a dummy variable equals one for the period after 2012. *COCA* is a dummy variable equal to one for municipalities in which their area suitable for coca cultivation is larger than the national average. All regression includes the following controls: sex, grade (in log), age, and age squared. Robust standard errors area clustered at the household level. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. P-values for standard errors in parenthesis: * significant at 10%, **significant at 5% and ***significant at 1%.

Two main findings stem from the estimation results. First, rural students seem more likely to be

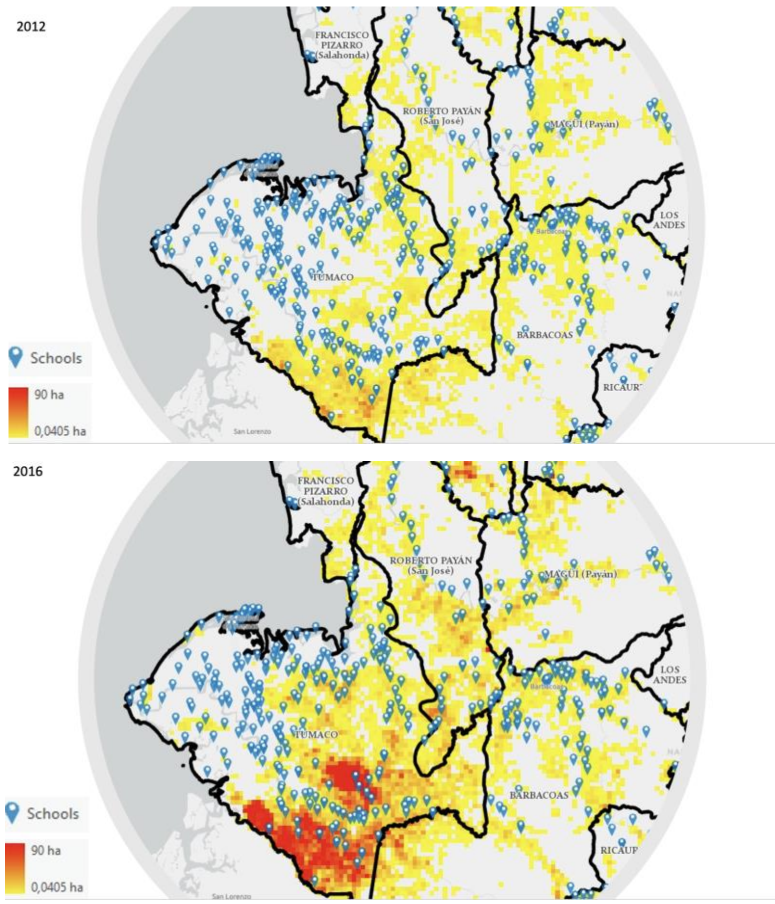
[51] According to Sviatchsi (2016), children may start by picking coca leaves, harvesting and drying them manually. Nevertheless, the relatively high earnings in the cocaine industry induce secondary children to participate in cocaine processing and transportation.

affected by changes in coca crops. Second, the de-escalation of the Colombian conflict may have induced children to drop out particularly in territories with high potential for coca crops. It may have occurred not only for the higher returns to working in coca crop-related activities but also for the reduction of violence that makes it less risky to work in these crops.

4.7 Mechanisms: Conflict De-escalation and Proximity to Coca Crops

Finally, to determine the correlation of the proximity to coca crops on dropout risk, we measure the effect of the distance to actual coca crops on the children and youth living in the rural areas of municipalities that exhibited above-the-mean exposure to FARC-EP violence before 2013. It is expected that the closer the school to coca crops, the higher the increase in the dropout risk. As shown in Figure 4.3, the large increase in coca crops during 2014-2016 was preceded by a decline in the number of hectares planted with coca crops that had started in 2005 and reached an all-time low in 2013 (UNODC, 2015). To estimate the effects of the expansion of coca crops on school dropout, we proceed as follows: (i) we geocode the locations of both of schools and coca crops during 2014-2016. Figure 4.6 depicts the large growth of coca crops in the most southwestern Colombia municipalities. It is apparent how schools located in *Tumaco* and *Rodrigo Payan* became heavily surrounded by coca crops; and (ii) we estimate the effect of the distance from school to the presence of coca crops on dropout risk using discrete measures of distance, as presented in Table 4.6.

Figure 4.6: Coca Crops near Schools Before and After 2013



Notes: This figure presents two maps of Tumaco and other municipalities. Tumaco is the municipality with the highest area of coca crops and was historically exposed to FARC-EP violence. The upper map corresponds to the year 2012 before the surge of coca crops, and the lower map corresponds to the year 2016. The yellow to red grid cells locate the coca crops and their density. The blue dots locate the rural schools. Source: UN-SIMCI and Colombian Ministry of Education.

Restricting the sample to only students from rural households in municipalities that experience above-the mean FARC-EP violence before 2013, we determine the effect of the distance to a coca crop on dropout using the following model:

$$y_{ihst} = \beta_0 + + \ \Sigma\beta_{j1} coca\ distance_{ihst} + \alpha X_{it} + \delta_h + \mu_t + \varepsilon_{ihmt} \qquad [4.4]$$

Building on our baseline equation, y_{ihmt} denotes the dropout event of student i, from household h, attending school *s* in year t (2014–2016). We define $coca\ distance_{ihst}$ as five dummy variables for five distance ranges of kilometers from the school (0-2, 2-5, 5-10, 10-20, more than 20) to the nearest coca crop. β_{j1} denotes the effect on dropout of the *j* = {0,…,5} ranges of distance.

Table 4.6: The Effect of Proximity to Coca Crops on School Dropout Risk

Dependent variable: Dropping out of school

	Municipalities highly exposed to FARC-EP violence, suitable for coca growing and rural household	
	After the unilateral ceasefire [1]	Before the unilateral ceasefire [2]
0-2 Km there is a coca crop from the school	0.0317***	-0.0156***
	(0.00720)	(0.00348)
2-5 Km there is a coca crop from the school	0.0160**	-0.00879***
	(0.00650)	(0.00295)
5-10 Km there is a coca crop from the school	0.00538	-0.00130
	(0.00528)	(0.00236)
10-20 Km there is a coca crop from the school	0.00787*	-0.00147
	(0.00405)	(0.00156)
Observations	178,790	1,377,436
Students	55059	302747
Households	31213	138403
R-squared	0.406	0.226
Households FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes

Notes: This table presents the results for the subsample of students from rural households in municipalities with potential suitable areas for coca cultivation higher than the national average and that experience above-the mean FARC-EP violence before 2013. We used five dummy variables for five distance ranges of kilometers from the school (0-2, 2-5, 5-10, 10-20, more than 20) to the nearest coca crop. The first column presents the results for the year after 2013 and the surge in coca crops, while the second column presents the results before 2013. All regression includes sex, grade (in log), age, and age squared as controls. Robust standard errors area clustered at the household level. p-values for standard errors in parenthesis: * significant at 10%, ** significant at 5% and *** significant at 1%.

Table 4.6 presents the estimations for β_{j1}. It is observed in column (1) students from schools up to two kilometers to the closest coca crop exhibit a dropout risk 3.1 percentage points higher than schools located more than 20 kilometers away from the nearest coca crop. For students from schools between two to five kilometers to the closest coca crop, the dropout risk is 1.6 percentage points higher. For schools located more than 10 kilometers away from a coca crop, the effect is smaller and not statistically significant. Column (2) in Table 4.6 reports the same coefficients for the period 2009–2012 during which coca crops seemed to be shrinking. As expected, the β_{j1} coefficients become negative, suggesting that the dropout risk diminished in areas of potential presence of coca crops.

4.8 Conclusions

This paper provides evidence that de-escalation of armed confrontations is not necessarily accompanied by improvements in the education of the population heavily affected by conflict. Counter-intuitively, the results obtained reveal that the de-escalation of the armed conflict in Colombia has negatively impacted school attendance of the most vulnerable students living in the municipalities historically affected by the actions of the FARC-EP. This study sheds light on the puzzling reality that face civilians living in conflict zones, especially when illicit activities are present. In a scenario where the end of hostilities takes place with the presence of resources highly susceptible to illicit exploitation, such as coca crops, short-term returns in those labor markets end up being highly attractive and incentivizing school dropout.

Our main results reveal that the risk of dropout for children and youth from the most vulnerable households increased by 13 percent compared to the control municipalities. These results are concentrated among children and youth living in the rural areas of the municipalities, with a 24 percent increased risk of dropping out compared to the control municipalities, while for children and youth living in urban areas, no such effect was found. As the main mechanism of this result, we identified the swift increase in illicit crops, which coincided with the de-escalation of the conflict in the municipalities historically most affected by violence in Colombia. For rural areas, where illicit crops are located and where the most vulnerable populations also live, the effect of de-escalation of conflict on the risk of dropout is fully explained by the presence of illicit crops. In contrast, for students living in urban areas, de-escalation of conflict is associated with a reduction in the dropout risk. Additionally, we found that the positive effect of de-escalation of violence on dropout risk is especially high for secondary school students. Older children and youth may be more attracted to performing low-skilled tasks in the cultivation of coca crops.

This study contributes to improve understanding of the relationship between armed conflict, illicit economies, and human development trajectories. Although the decrease in hostilities facilitates the provision of public goods and access to them, de-escalation scenarios with a territorial predominance of illicit activities may end up having an unintended impact on the human development trajectories of the most vulnerable inhabitants.

Chapter 5. The impact of ICT use at school on student achievement: causal evidence from PISA based on variability within student[£]

5.1 Introduction

The use of ICT for learning at school has boomed during the last decade. Governments across the world, assuming that ICT use would have a positive impact on student achievement, have made massive investments in equipping schools with ICT devices and tools. The COVID-19 pandemic accelerated even further the incorporation of ICT into education. However, existent research evidence on the causal impact of ICT use at school on student achievement is uneven (Spiezia, 2010; OECD, 2015; Fernández-Gutiérrez et al., 2020). The massive introduction of ICT in education has often been based on optimistic preconceptions among policy-makers (and, in recent years, on imperative responses to urgent needs posed by the pandemic), rather than on rigorous academic evidence able to prove benefits of ICT use for learning.

ICT use for learning may render benefits on student achievement, resulting from diverse factors (Fernández-Gutiérrez et al., 2020): 1) An increase in the amount of information and resources available for learning; 2) A favourable impact of ICT on the attractiveness or interactivity of lessons; 3) An increase of students' flexibility and autonomy; and 4) Facilitation of individualized instruction and monitoring of student progress. On contrary, ICT use for learning may also have negative impacts on students' outcomes, derived from a series of undesired outcomes which may be favoured by the use of ICT tools: 1) Distraction of students; 2) Undermine of students' need for work and discipline; 3) Restriction of students' creativity; and 4) Reduction of students' real interaction with teachers. It is unclear whether, and in which contexts, the positive or negative

[£] This chapter has as coauthors Marcos Fernández-Gutiérrez (Universidad de Cantabria) and Gregorio Gimenez (Universidad de Zaragoza).

effects of ICT use for learning on student achievement prevail (Falck et al., 2018; Vargas-Montoya et al., 2023).

There is a growing body of research focused on analysing the impact of ICT use at school on student achievement. This literature can be classified into two groups. On the one hand, experimental and quasi-experimental studies, based on research designs which allow inferring a causal impact of ICT use on student achievement. On the other hand, analyses based on large-scale international surveys on students' cognitive outcomes (e.g.: PISA, PIRLS and TIMMS), which provide evidence representative at the student population level and comparable across countries. While both approaches have strengths, both have also significant shortcomings. Experimental and quasi-experimental studies allow finding causal effects but, as they are usually based on ad-hoc studies which use specific measures of ICT use and student achievement, it is difficult to generalize their results and even to compare them across different contexts and countries (De Witte and Rogge, 2014). On contrary, whilst studies based on large-scale international surveys solve this issue, they typically face a problem of endogeneity (Fariña et al., 2015): non-observable factors (such as students' intelligence or motivation) may be simultaneously associated with students' use of ICT and student achievement. For this reason, the results obtained by these studies should be interpreted as statistical correlations between ICT use and student achievement, rather than as a causal impact of ICT use on student achievement.

This paper contributes to this literature by estimating the causal impact of both ICT use for learning at school and the intensity of ICT use for learning at school on students' outcomes in PISA (the most well-known international survey on student achievement). To do so, using data for 135,000 students from 35 countries obtained from PISA-2018, the paper takes advantage of the within-student variability in student achievement and ICT use across two different subjects: mathematics and science. We follow a two steps approach. First, for the whole sample of 35 countries we estimate the association between both ICT use and the intensity of ICT use for learning at school and students' outcomes in PISA. Then, after testing the necessary conditions for a causal interpretation of the results, we identify the impact of both ICT use and the intensity of ICT use for learning at school on students' outcomes for those countries in which these conditions hold. The results obtained show that the impact of both ICT use and the intensity of ICT use for learning at school on students' outcomes in PISA varies depending on the country, albeit it is non-significant in most countries.

The rest of the paper is structured as follows. After this introduction, the second section describes the previous literature that addressed the relationship between ICT use at school and student achievement. The third section explains the empirical strategy followed by the paper. The fourth section describes the data used. The fifth section summarizes the results obtained. The sixth section concludes.

5.2 Literature review

ICT use at school and student achievement: evidence from experimental and quasi-experimental analyses

In the last two decades, some papers have estimated the impact of ICT use at school on student achievement from experimental or quasi-experimental approaches. Each of these studies was carried out for a specific country and context and used specific measures of both ICT use and student achievement. A majority of these studies based on experimental or quasi-experimental approaches found non-significant effects of ICT use on student achievement.

Angrist and Lavy (2002) estimated the effect of a large-scale program for installing computers at some schools in Israel (configured as a natural experiment) on student achievement in math and language. Student achievement was measured by scores in tests carried out by a national institution among 4th grade and 8th grade students. These authors did not find an improvement in students' scores in those schools where computers were installed (even, they found a negative effect on math score among 4th grade students).

Rouse and Krueger (2004) carried out an experimental evaluation of the use of a computer program for learning language and reading among students with learning difficulties in four US schools. They found non-significant effects of the use of the program on students' language and reading skills, measured according to standardized tests developed at the State level. In contrast, these authors found that the program improved students' language skills if they were measured according to tests developed by the company which produced the program.

Goolsbee and Guryan (2006) evaluated a subsidy on introducing the Internet in public schools in California, using a regression discontinuity design that exploits variation in subsidy rates across schools. They found that, although the subsidy increased investment in the Internet, it did not have significant effects on student achievement, measured by scores in a variety of tests on math, reading and science skills adapted from a standard test used at the State level.

Craig et al. (2013) carried out an experiment for evaluating the introduction of an intelligent tutoring system for after-school learning of math in four US schools. They obtain that this tool had non-significant effects on student achievement, measured by scores in a standard test used at the State level.

Cristia et al. (2017), from other experiment, evaluated the effect of a program which provided laptops to rural students in Peru, to be used both at home and at school. These authors found that, although the program increased the use of computers, it had non-significant effects on student achievement in math and language, measured by scores in tests specifically prepared for this research (which were based on standard national examinations). In contrast, they found positive effects of the program on scores in tests which measured cognitive skills.

Some experimental studies, however, have obtained significant effects of ICT use at school on student achievement, in certain specific settings. Some of these studies found a positive impact of ICT use on student achievement. In particular, positive effects have been found in some specific settings in which ICT are used among students in disadvantaged positions.

Banerjee et al. (2007), from two experiments, evaluated the implementation of a computer-assisted program for learning math among students from poor families in urban schools in India. These authors found that the program increased student achievement in its first and second year of implementation, measured by scores in specific tests designed to cover basic competences taught in grades 1 to 4. They attribute the result to the wide social distance between students and their teachers in the specific context where the experiments were carried out, which hinders interaction between students and teachers when teaching is not mediated by ICT. As a result, the potential negative effects of introducing ICT which are associated with a reduction of that interaction would be limited.

Barrow et al. (2009), also from an experiment, evaluated the use of a computer program for learning algebra in three US urban school districts with a high proportion of students from ethnic minorities. These authors obtained that the program improved student achievement, measured by a variety of tests developed by an independent association, by States and by districts. They found that, however, this effect was not observed in all of the districts, and it was higher where classes were larger and more heterogeneous, and where absenteeism was higher. They explain these results highlighting the effect of the computer program on increasing individual attention of teachers to students which required additional assistance.

Bartelet et al. (2016), from another experiment, evaluated a web-based tutoring system for learning math offered for homework by a school in the Netherlands. They found a positive effect of this tutoring system on student achievement, measured by scores in tests designed by the company which develops the tool. This effect depended on the math domain considered and was higher among low achievers.

Leuven et al. (2007), in contrast, found a negative impact of ICT use on student achievement in a different setting. These authors used a regression discontinuity design to evaluate a subsidy on the acquisition of computers and software among schools in the Netherlands, targeted at schools with a high proportion of students from ethnic minorities or from families with a low educational attainment. They found a negative effect of the subsidy on student achievement in language and math, measured by scores in national tests among students in 8th grade.

In general, a majority of experimental and quasi-experimental studies have found a non-significant impact of ICT use at school on student achievement. Some studies, however, have found a positive, or a negative, impact of ICT use on student achievement in certain settings. Experimental and quasi-experimental are generally carried out in specific settings and contexts, which hinders their external validity. In addition, they use specific measures of student achievement, generally from tests available in the specific setting in which the study is carried out and evaluate specific programs introducing a particular ICT at schools. All of this hinders the comparability of these studies.

ICT use at school and student achievement: evidence from international surveys

In recent years, the availability of information on both student achievement and ICT use from large-scale international surveys has favoured the development of a number of papers which exploited these surveys to estimate the relationship between ICT use and students' scores in the surveys. As explained above, as these analyses use standard measures of both student achievement and ICT use available in these surveys, they permit their replicability and their comparability across different settings and countries. However, the presence of non-observable factors at the student level which may be simultaneously associated with both student achievement and ICT use hinders that the statistical relationship obtained between both issues can be interpreted as a causal impact of ICT use on student achievement (Spiezia, 2010; Fariña et al., 2015).

The rich amount of information available in international surveys (PIRLS, TIMMS and, in particular, PISA) have permitted flourishing of analyses on the relationship between student achievement and both ICT availability and ICT use. Information in these surveys include both ICT availability and ICT use, and both at home and at school. Given the purpose of our study, and that the problem of endogeneity may be particularly strong for ICT availability and use at home, we restrict our literature review on the analyses focused on ICT availability and use at school only.

The studies analyzing the relationship between ICT and student achievement from international surveys have found, in general, a positive relationship between ICT availability at school and students' outcomes measured in the tests, and a negative relationship between ICT use at school and students' outcomes in those tests. Among studies made for a single country, Mediavilla and Escardíbul (2015), using PISA-2012 data for Spain, obtained a positive relationship between ICT availability at school and PISA scores in math, albeit not in reading. In addition, these authors found a negative relationship between ICT use at school (measured by the frequency of use of a computer for a series of activities) and PISA scores in math (and among boys, also for scores in reading). They also found that the intensity of ICT use (measured by the amount of time using computers) was negatively related to PISA scores among girls. Also using PISA-2012 data for Spain, Alderete et al. (2017) analysed the relationship between both access to and use of ICT at school and PISA scores in math, reading and science, by estimating a structural equation model (where the relationship between ICT access and student achievement is mediated by ICT use). From this approach, these authors found that both access to and use of ICT at school are negatively related to PISA scores in the three subjects. In other paper, Erdogdu and Erdogdu (2015) used PISA-2012 data for Turkey to analyse the relationship between availability and use of internet and PISA scores. They found that the availability of internet connection at school was positively related to PISA scores in science, whilst the frequency of browsing the internet at school was negatively related to PISA scores in math, science and reading.

Other studies have exploited the availability of homogeneous information across countries in international surveys for carrying out analyses of the relationship between ICT use at school and student achievement for a set of countries. Zhang and Liu (2016) used data from five waves of PISA (from 2000 to 2012), for all the countries where information was available (from 25 countries in PISA-2000 to 43 countries in PISA-2012), for analysing the relationship between student variables on both ICT availability at school and ICT use at school and student achievement in math and science. These authors found that ICT availability at school was positively related to PISA scores, whilst ICT software and the use of the Internet were negatively related to these scores. Petko et al. (2017), using PISA-2012 data for 39 countries, also found that ICT use at school was negatively related to PISA scores in math, reading and science. Skryabin et al. (2015), using data from multiple international surveys (TIMMS-2011, PIRLS-2011 and PISA-2012) for all the countries for which information was available in each survey (43, 38 and 39 countries, respectively), found that the relationship between ICT use at school and student achievement differed by students' grade: among students in 8th grade, ICT use at school was negatively related to test scores, whilst among students in 4th grade they observed a positive relation between both issues. In another study, Hu et al. (2018), using PISA-2015 data for 44 countries, analyzed the relationship between both ICT availability and

ICT use at school and student achievement controlling for students' interest, competence, autonomy and enjoyment of social interaction when using ICT. These authors found that, after controlling for these factors, the relation between ICT availability at school and students' scores in PISA was non-significant, whilst the relation between ICT use and PISA scores kept negative.

Some papers which analyzed the relationship between ICT use at school and student achievement from international surveys have explored specific research designs to address endogeneity issues. Spiezia (2010), in a study based on PISA-2006 data for the 33 countries where information was available, analyzed the relationship between computer use and students' scores in PISA controlling for students' observable characteristics and self-selection. To do so, this author first estimated the frequency of computer use as a function of students', their families' and their schools' observable characteristics. Next, he estimated students' scores in PISA as a function of those observable characteristics, the frequency of computer use and the residuals from the first estimation (as a measure of unobserved students' characteristics). From this approach, he obtained a non-significant effect of computer use at school on PISA scores in most countries.

De Witte and Rogge (2014), using TIMMS-2011 data for the Netherlands, estimated the effect of ICT use on student achievement by applying matching techniques. This approach allows to account for observable characteristics of students, teaching, administrative personnel and school management. These authors conclude that, after taking into account these variables, the effect of ICT use became non-significant.

Cabras and Tena Horrillo (2016), using PISA-2012 data for Spain, analysed the effect of investment in ICT on student achievement in math by applying a non-parametric method: Bayesian Regression Trees (BART). They found a positive effect of ICT use on students' PISA scores in this subject, particularly among students from a low socioeconomic background. The same method was applied by Ferraro (2018), using PISA-2012 data for Italy, for analyzing the effects of ICT use at school on student achievement in math. She also found a positive effect of ICT use on students' PISA scores in this subject.

Falck et al. (2018) took TIMMS-2011 data for 30 countries to analyze the effect of using computers at school lessons on student achievement in math and science. To do so, they exploited within-student between-subject variability in computer use, derived from availability of information on two different subjects (math and science) for the same students, which allows to control for unobservable factors at the student level. A similar approach had been applied by Comi et al. (2017) for estimating the effect of ICT use at school on student achievement in math and language, using data from two surveys administered to a sample of students and their teachers in Lombardy, combined with administrative data on their scores in a national test. Both Comi et al. (2017) and Falck et al. (2018) concluded that the effect of ICT use on student achievement depended on the type of use of ICT. In particular, Falck et al. (2018) found that the use of computers to practice skills and procedures had a positive effect on students' scores in TIMMS, whilst the use of computers to process and analyze data had non-significant effects.

Finally, Fernández-Gutiérrez et al. (2020) combined data from three rounds of PISA (2009, 2012 and 2015) for Spain to analyze the effect of changes in ICT use at school in Spanish regions on student achievement. This analysis exploited the autonomy of Spanish regions, and variability across them in decisions on ICT use at schools, as well as representativeness of PISA data in terms of the student population for each Spanish region. They obtained that the increase of ICT use at school had non-significant effects on PISA scores in math and reading, albeit it has a positive effect on PISA scores in science.

In sum, analyses based on data from international surveys show that the estimated effect of ICT use at school on student achievement varies depending on the empirical approach adopted. Without accounting for endogeneity, the relationship between variables measuring ICT use and students' outcomes in tests provided by these surveys tends to be negative. However, when specific methods for controlling for students' unobservable characteristics are applied, the relationship between variables on ICT use and students' outcomes tend to become non-significant, and even positive in certain cases.

5.3 Empirical strategy

Evaluating the effect of ICT use on student achievement faces multiple challenges in terms of identification. As explained above, the existence of unobservable factors at both the student and the school level (for example, students' and teachers' motivation in ICT use, or how and when the students began to use ICT) and their correlation with both ICT use and academic performance imply the need to seek an approach that allows controlling for such factors. One possibility to control for such unobservable factors at the individual and the school level is to exploit within-student variation on student achievement and specific use of ICT through different subjects. Equation [5.1] presented below describes the empirical approach adopted in this paper to exploit within-student variation from the information available in PISA:

$$y_{iks} = \mu_i + \delta ICT_{ks} + \boldsymbol{\pi X_{ik}} + \boldsymbol{\varphi S_k} + \gamma_s + \vartheta_k + \epsilon_{isk} \qquad [5.1]$$

Where y_{iks} corresponds to the performance of student *i* in school *k* in subject *s*, ICT_{ks} is ICT use in school *k* in subject *s*, X is a vector of characteristics of student *i* in school *k*, S is a vector of characteristics of school *k*. μ_i corresponds to a fixed effect at the individual level, γ_s to a fixed effect at the subject level and ϑ_k to a fixed effect at the school level. ϵ_{isk} is the unexplained component. By exploiting the within-student variation, all those characteristics (observable and unobservable) of the student and the school context and their respective interactions (which are common to the student and therefore across subjects) are captured.

This approach controls elements such as the socioeconomic background of the student, their abilities and preferences, the quality of the school, and the educational infrastructure in terms of technology and connectivity. However, this technique entails a strong identification assumption: that the effect of ICT use on student achievement is the same across the different subjects. To test this, we make use of the proposal of Metzler and Woessmann (2012), who employed data from the 2004 Peruvian national evaluation of student achievement. We adapt it to the math and science subjects of the PISA dataset, by modelling the unobservable fixed effect at the individual level through equation [5.2].

$$\mu_i = \alpha_{mathk}\, ICT_{mathk} + \alpha_{sciencek}\, ICT_{sciencek} + \omega X_k + \rho S_k + \varepsilon_i \quad [5.2]$$

Then by substituting [5.2] in [5.1] and rearranging terms we have the following:

$$y_{iks} = (\delta_s + \alpha_s) ICT_{ks} + \sum_{t \neq s} \alpha_t ICT_{tk} + (\pi + \omega) X_{ik} + (\varphi + \rho) S_k + \gamma_s + \vartheta_k + \varepsilon_i + \epsilon_{isk} \qquad [5.3]$$

The estimation of the system of equations in [5.3] allows us to test together two central elements of the assumption: (1) $\alpha_{math} = \alpha_{science}$, that is, if the relationship between the ICT use and the unobservable fixed effect is the same for all the subjects; and (2) $\delta_{math} = \delta_{science}$, that is, if ICT use has the same impact on student achievement for both subjects.

5.4 Data

In order to estimate equations [5.1] and [5.3], this paper uses data from PISA-2018. The OECD Program for International Student Assessment (PISA) is the most well-known large scale international survey which evaluates students' knowledge and skills. It was launched for the first time in 2000 and is carried out every 3 years in different countries around the world since then. PISA target population in every country is composed of students aged between 15 years and three months to 16 years and two months enrolled in educational institutions in grade 7 or higher. The random sampling process adopted in PISA ensures the representation of the entire target population at the country level (OECD, 2020).

In PISA, students take tests in mathematics, reading and science. Given its international nature, PISA does not focus on a specific curriculum, but rather on how students can use their knowledge in real-life problems and situations. Additionally, students provide background information about their families, the educational environment of their homes, and the use of tools for academic purposes. Principals also participate reporting schools' characteristics.

This paper takes the information from PISA-2018 as a starting point, a large international sample of 612,004 students attending 21,903 schools in 79 countries and territories.[52] We focus on students from those 35 countries which fulfilled two conditions: (a) Students were tested in mathematics and science and reported having these subjects as part of their classes (b) Students answered questions regarding the ICT use in academic activities inside the schools (included in a specific questionnaire only used in some countries).

[52] For simplicity, we refer to all of them throughout this paper as countries regardless of their administrative status.

The first condition is established in search of the empirical approximation around the impact of ICT on student achievement to focus on a couple of subjects that may have more affinity from the need and availability of ICT tools for their learning, in this case, *math* and *science*. The second condition is linked to a characteristic that is particular to PISA-2018 and essential for applying the empirical approach used in this paper. In contrast to previous rounds of PISA, PISA-2018 included a set of questions to measure ICT use disaggregated by subject. The use of this information on subject-specific ICT use for learning at school constitutes a key novelty for our study because it allows us to exploit within-student variation to identify the impact of ICT use on student achievement. With the limitations imposed by these two conditions, the final sample we work with is reduced to 135,000 students in 35 countries. The sample, which includes countries of America (8), Asia (9) and Europe (18), is widely heterogeneous in terms of countries' population size, economic characteristics, academic performance and educational system characteristics.

To estimate equations [5.1] and [5.3], PISA scores in science and mathematics were standardized to a distribution with mean zero and variance 1. To measure ICT use and intensity of ICT use for learning at school, this paper adopts two approaches. The first approach, focused on measuring ICT use for learning at school, takes as its starting point PISA question IC152 *(Within the last month, has a digital device been used for learning or teaching during lessons in the following subjects?*). For each student and subject, this variable takes the value of 1, if for the respective subject it is reported any of the following options: "Yes, both the teacher and students used it*"* or "Yes, but only students used it*"*, otherwise it takes the value of 0. The second approximation takes as its starting point PISA question IC150 *(In a typical school week, how much time do you spend using digital devices during classroom lessons?*), and it is meant to approximate the intensity of ICT use for learning at school, as the number of minutes of student use during a typical week. We follow the approximation proposed by Lavy (2015) to measure differences in schools' instruction time and assign to each student and subject the average value of the time interval of use reported by the student (0 minutes, 15 minutes, 45 minutes and 60 minutes).

For the estimation of the system of equations in [5.3], gender, age, PISA index of economic, social and cultural status, sense of belonging to the school and subjective well-being are included as part of the vector of student characteristics (X). Regarding the group of variables at the school level (S), it includes the proportion of certified teachers, the index of obstacles to learning by teachers, the perceived interest of teachers, the percentage of financing coming from the government, the adaptability of instruction, report on short staff, and characteristics of the community where the school is located. Although the estimation of equation [5.1] by exploiting within-student variation considers all student and school level variables that are fixed across subjects, the system of equations [5.3] requires the explicit inclusion of student and school level predictors of the academic performance of the students in each one of the subjects. To pick these variables, we consider predictors identified by other studies with standardized tests performance (Lee and Stankov, 2018; Gamazo and Martinez-Abad, 2020; OECD, 2020)

5.5 Results

Tables 5.1 and 5.2 present, respectively, the estimates of the effect of ICT use for learning at school on student achievement (corresponding to equation [5.1]) and the results from the joint test of the restrictions α_{math} = $\alpha_{science}$ and δ_{math}= $\delta_{science}$ (derived from the system of equations in [5.3]), with ICT use for learning at school in each subject measured from PISA question IC152 (use of digital devices by students and teachers). Columns (1) to (3) in Table 5.1 show the value of the estimator for δ, standard error, and t statistic corresponding to equation [5.1] for each of the 35 countries included in the sample, restricting to observations for which information on all the student level controls used in the estimation of the system of equations [5.3] is available. Likewise, Columns (4) to (6) in Table 5.1 show the value of the estimator for δ, standard error and t statistic corresponding to equation [5.1] for each of the 35 countries included in the sample, considering those observations for which information on all the controls at both the student and school level used in the estimation of the system of equations [5.3] is available.

Tables 5.3 and 5.4 present, respectively, the estimates of the effect of the intensity of ICT use for learning at school on student achievement (corresponding to equation [5.1]) and the results from the joint test of the restrictions α_{math} = $\alpha_{science}$ and δ_{math}= $\delta_{science}$ (derived from the system of equations in [5.3]), measuring the intensity of ICT use for learning at school in each subject from PISA question IC150. In Table 5.3, Columns (1) to (3) show the value of the δ estimator, standard error, and t-statistic corresponding to equation [5.1]. They are calculated for each of the 35 countries in the sample, using the student level controls included the system of equations [5.3]. Likewise, Columns (4) to (6) show the δ estimator, standard error, and t-statistic corresponding to equation [5.1] for the 35 countries, including controls at both the student and school level used in the estimation of the system of equations [5.3].[53] The results are presented divided into two groups in order to see if the inclusion of controls at the school level, in addition to those at the student level, has an effect on the point estimators. The correlation of the estimators of columns (1) and (4) in Table 5.1 is 0.97, whilst that in Table 5.3 is 0.84, which indicates that the results are quite robust across the different specifications.

The results employing ICT use for learning at school as a variable of interest (Tables 5.1 and 5.2) reveal that of the 35 countries under analysis, the great majority (between 27 and 30 depending on

[53] Tables D.1, D.2, D.3 and D.4 in the annexes present results under the same structure as this section but using mathematics and reading. Although less affinity between mathematics and language is expected as far as the use of ICT is concerned, these results are included in order to show the applicability of the methodology used.

the specification used) present a relationship between ICT use for learning at school and students' outcomes that is not statistically different from zero when the within-student variation is exploited. Among those with a relationship statistically different from zero, we observe that Costa Rica, Croatia, Germany, Malta and Panama show a positive relationship between ICT use for learning at school and students' outcomes; while Hong Kong, Iceland, Kazakhstan, Serbia and Tatarstan (Russia) show a negative relationship between both issues. However, as we have explained previously in this paper, to interpret the results as a causal relation, the restrictions derived from the equation system in [5.3] have to be met. Table 5.2 reveals that only 5 countries (the Dominican Republic, Georgia, Iceland, Malta, and Mexico) meet the necessary conditions. Of these, Dominican Republic, Georgia, and Mexico show effects of ICT use for learning at school on students' outcomes that are not statistically different from zero. Iceland presents a negative effect. In contrast, Malta presents a positive effect.

The results considering the intensity of ICT use for learning at school (in terms of the amount of time of use of ICT) as a variable of interest (Tables 5.3 and 5.4) reveal that of the 35 countries, the great majority (between 29 and 30 depending on the specification used) present also a relationship between the intensity of ICT use for learning at school and students' outcomes that is not statistically different of zero when the within-student variation is exploited. Among those with a relationship statistically different from zero, we observe that in Finland, France, Korea, Malta, and the Slovak Republic the intensity of ICT use for learning at school has a positive relationship with students' outcomes; while in Brazil, Kazakhstan, and Serbia the relationship is negative. However, as previously mentioned, to interpret the results as a causal relation, the restrictions derived from the equation system in [5.3] have to be met. Table 5.4 reveals that only Brazil, Iceland, Malta, and Poland meet the necessary conditions. Of these, Iceland and Poland show effects of intensity of ICT use for learning at school on students' outcomes that are not statistically different from zero. Brazil presents a negative effect. In contrast, Malta presents a positive effect.

For the complete sample of 35 countries, the association between both ICT use for learning at school and intensity of ICT use for learning at school and student achievement shows heterogeneous results across countries. An association not statistically different from 0 is the most predominant, observed for a number between 27 and 30 of the countries considered, depending on the specification. The results reveal that among those countries that meet the necessary condition for a causal interpretation of this relationship derived from the within-student variation, both ICT use and intensity of ICT use for learning at school show mixed effects on student achievement across countries. In both cases, a non-significant effect of ICT on student achievement is the result most frequently obtained.

Table 5.1: ICT use and student achievement

Country or territory	Observations with non-missing student controls			Observations with non-missing student & school controls		
	(1)	(2)	(3)	(4)	(5)	(6)
	δ	**se**	**t**	**δ**	**se**	**t**
Albania	0,127	0,089	1,435	0,123	0,091	1,355
Brazil	0,001	0,086	0,011	0,043	0,095	0,449
Bulgaria	-0,154	0,115	-1,330	-0,130	0,134	-0,973
Chile	0,115	0,073	1,572	0,155	0,081	1,920
Chinese Taipei	0,056	0,074	0,752	0,071	0,077	0,922
Costa Rica	0,134	0,062	2,180	0,132	0,062	2,128
Croatia	0,153	0,069	2,233	0,169	0,073	2,313
Czech Republic	0,065	0,063	1,029	0,057	0,065	0,882
Dominican Rep.	0,136	0,134	1,014	0,218	0,156	1,398
Estonia	-0,023	0,059	-0,388	-0,023	0,059	-0,393
Finland	0,048	0,054	0,884	0,054	0,056	0,962
France	0,057	0,037	1,555	0,063	0,044	1,435
Georgia	0,071	0,089	0,806	0,098	0,094	1,039
Germany	0,140	0,073	1,932	0,230	0,088	2,613
Greece	-0,044	0,072	-0,612	0,008	0,081	0,099
Hong Kong	-0,172	0,101	-1,695	-0,246	0,120	-2,043
Iceland	-0,094	0,045	-2,102	-0,114	0,053	-2,161
Japan	0,006	0,066	0,087	0,008	0,067	0,117
Kazakhstan	-0,266	0,128	-2,072	-0,270	0,130	-2,072
Korea	0,037	0,048	0,781	0,062	0,054	1,153
Lithuania	-0,059	0,091	-0,642	-0,058	0,093	-0,622
Luxembourg	0,114	0,099	1,150	0,230	0,161	1,429
Malta	0,276	0,139	1,991	0,299	0,152	1,972
Mexico	0,126	0,094	1,342	0,082	0,111	0,739
Panama	0,258	0,109	2,366	0,355	0,148	2,396
Poland	0,011	0,111	0,098	0,011	0,111	0,100
Russian Federation	0,027	0,106	0,254	0,019	0,108	0,172
Serbia	-0,217	0,120	-1,806	-0,260	0,124	-2,102
Slovak Republic	0,097	0,064	1,521	0,042	0,065	0,647
Switzerland	-0,110	0,081	-1,361	-0,107	0,075	-1,430
Thailand	0,120	0,066	1,819	0,116	0,066	1,761
Turkey	-0,008	0,063	-0,132	0,006	0,064	0,099
United Kingdom	0,036	0,040	0,922	0,015	0,056	0,274
United States	-0,023	0,064	-0,365	-0,042	0,071	-0,596
Uruguay	-0,083	0,084	-0,997	-0,099	0,083	-1,200
Moscow region	0,066	0,172	0,383	0,052	0,178	0,293
Tatarstan (rus)	-0,209	0,090	-2,320	-0,194	0,091	-2,123

Table 5.2: ICT use and student achievement necessary test

Country or territory	Student controls		Student and school controls	
	(1)	**(2)**	**(3)**	**(4)**
	Chi2	**Prob > chi2**	**Chi2**	**Prob > chi2**
Albania	124,435	0,000	91,891	0,000
Brazil	83,848	0,000	8,421	0,015
Bulgaria	32,603	0,000	36,184	0,000
Chile	61,087	0,000	48,803	0,000
Chinese Taipei	203,324	0,000	132,893	0,000
Costa Rica	46,586	0,000	22,453	0,000
Croatia	1119,647	0,000	986,342	0,000
Czech Republic	301,609	0,000	196,417	0,000
Dominican Republic	9,865	0,007	3,725	0,155
Estonia	38,016	0,000	44,485	0,000
Finland	9,635	0,008	7,312	0,026
France	1435,181	0,000	842,008	0,000
Georgia	0,474	0,789	2,676	0,262
Germany	467,049	0,000	330,122	0,000
Greece	209,447	0,000	263,722	0,000
Hong Kong	396,409	0,000	139,625	0,000
Iceland	4,421	0,110	6,226	0,044
Japan	13,870	0,001	38,255	0,000
Kazakhstan	1401,537	0,000	1123,045	0,000
Korea	135,578	0,000	58,309	0,000
Lithuania	321,806	0,000	193,135	0,000
Luxembourg	88,779	0,000	11,018	0,004
Malta	36,785	0,000	3,814	0,149
Mexico	5,811	0,055	3,451	0,178
Panama	218,550	0,000	133,745	0,000
Poland	37,115	0,000	33,515	0,000
Russian Federation	234,531	0,000	151,640	0,000
Serbia	117,883	0,000	171,831	0,000
Slovak Republic	314,491	0,000	180,504	0,000
Switzerland	198,404	0,000	114,218	0,000
Thailand	603,989	0,000	445,089	0,000
Turkey	257,832	0,000	128,437	0,000
United Kingdom	44,301	0,000	10,600	0,005
United States	77,514	0,000	39,325	0,000
Uruguay	41,020	0,000	26,535	0,000
Moscow region	16,930	0,000	16,929	0,000
Tatarstan (rus)	52,283	0,000	68,066	0,000

Table 5.3: The time of use of ICT and student achievement

Country or territory	Observations with non-missing student controls			Observations with non-missing student & school controls		
	(1)	(2)	(3)	(4)	(5)	(6)
	δ	**se**	**t**	**δ**	**se**	**t**
Albania	0,0007	0,0007	0,9454	0,0005	0,0008	0,7041
Brazil	-0,0011	0,0007	-1,5166	-0,0017	0,0009	-2,0186
Bulgaria	0,0006	0,0008	0,7633	0,0007	0,0008	0,8048
Chile	0,0005	0,0005	0,9635	0,0004	0,0006	0,6269
Chinese Taipei	-0,0003	0,0004	-0,7738	-0,0005	0,0004	-1,4760
Costa Rica	0,0004	0,0005	0,8212	0,0004	0,0005	0,8718
Croatia	0,0001	0,0005	0,1781	0,0003	0,0005	0,5574
Czech Republic	-0,0005	0,0005	-0,9289	-0,0003	0,0005	-0,5744
Dominican Republic	0,0006	0,0009	0,6604	0,0014	0,0013	1,1448
Estonia	0,0003	0,0005	0,5736	0,0003	0,0005	0,5826
Finland	0,0013	0,0005	2,6629	0,0011	0,0005	2,1921
France	0,0012	0,0004	3,2804	0,0010	0,0004	2,4122
Georgia	-0,0001	0,0008	-0,0663	-0,0003	0,0008	-0,4235
Germany	0,0009	0,0005	1,8068	0,0010	0,0006	1,6014
Greece	-0,0002	0,0005	-0,3952	-0,0003	0,0006	-0,4183
Hong Kong	0,0005	0,0012	0,4107	-0,0010	0,0007	-1,3779
Iceland	-0,0006	0,0005	-1,2253	-0,0006	0,0005	-1,1470
Japan	0,0001	0,0005	0,2191	0,0001	0,0005	0,2227
Kazakhstan	-0,0009	0,0005	-2,0080	-0,0009	0,0005	-1,9939
Korea	0,0010	0,0003	3,2969	0,0011	0,0003	3,2509
Lithuania	0,0005	0,0005	0,8576	0,0005	0,0005	0,8327
Luxembourg	-0,0001	0,0005	-0,1437	-0,0001	0,0007	-0,0842
Malta	0,0014	0,0007	1,9733	0,0015	0,0007	2,0656
Mexico	-0,0001	0,0006	-0,1392	-0,0005	0,0006	-0,8094
Panama	0,0023	0,0012	1,9189	0,0002	0,0011	0,1869
Poland	0,0005	0,0005	1,1237	0,0006	0,0005	1,1954
Russian Federation	-0,0002	0,0004	-0,5063	-0,0001	0,0004	-0,3124
Serbia	-0,0015	0,0007	-1,9992	-0,0014	0,0007	-1,8579
Slovak Republic	0,0016	0,0006	2,6605	0,0016	0,0006	2,6780
Switzerland	-0,0012	0,0006	-1,8310	-0,0008	0,0007	-1,2019
Thailand	0,0005	0,0005	0,8669	0,0004	0,0005	0,8100
Turkey	-0,0001	0,0003	-0,4530	-0,0001	0,0003	-0,3493
United Kingdom	0,0000	0,0005	0,0562	-0,0002	0,0008	-0,2169
United States	0,0002	0,0004	0,6633	0,0000	0,0004	-0,0735
Uruguay	0,0002	0,0007	0,2825	0,0002	0,0007	0,3532
Moscow region (rus)	0,0012	0,0007	1,6496	0,0012	0,0007	1,6820
Tatarstan (rus)	-0,0004	0,0005	-0,7670	-0,0003	0,0005	-0,6667

Table 5.4: The time of use of ICT and student achievement necessary test

Country or territory	Student controls		Student and school controls	
	(1)	(2)	(3)	(4)
	Chi2	**Prob > chi2**	**Chi2**	**Prob > chi2**
Albania	102,95	0,00	68,46	0,00
Brazil	66,20	0,00	4,00	0,14
Bulgaria	45,55	0,00	43,96	0,00
Chile	44,20	0,00	30,72	0,00
Chinese Taipei	140,45	0,00	67,55	0,00
Costa Rica	33,30	0,00	17,20	0,00
Croatia	973,09	0,00	889,94	0,00
Czech Republic	387,62	0,00	270,11	0,00
Dominican Republic	78,83	0,00	20,82	0,00
Estonia	23,87	0,00	27,35	0,00
Finland	19,06	0,00	12,16	0,00
France	1090,16	0,00	651,27	0,00
Georgia	1,90	0,39	8,89	0,01
Germany	462,78	0,00	358,63	0,00
Greece	192,25	0,00	144,32	0,00
Hong Kong	431,07	0,00	169,86	0,00
Iceland	1,03	0,60	0,22	0,90
Japan	14,29	0,00	36,53	0,00
Kazakhstan	956,20	0,00	726,87	0,00
Korea	194,70	0,00	103,03	0,00
Lithuania	276,37	0,00	154,54	0,00
Luxembourg	65,28	0,00	7,68	0,02
Malta	25,21	0,00	1,17	0,56
Mexico	18,84	0,00	5,80	0,06
Panama	124,47	0,00	70,44	0,00
Poland	6,62	0,04	3,41	0,18
Russian Federation	137,56	0,00	101,81	0,00
Serbia	26,73	0,00	42,31	0,00
Slovak Republic	172,15	0,00	69,89	0,00
Switzerland	151,15	0,00	110,37	0,00
Thailand	626,68	0,00	401,88	0,00
Turkey	290,10	0,00	174,89	0,00
United Kingdom	12,80	0,00	22,83	0,00
United States	48,49	0,00	22,69	0,00
Uruguay	15,01	0,00	9,86	0,01
Moscow region (rus)	14,00	0,00	15,60	0,00
Tatarstan (rus)	70,31	0,00	78,55	0,00

5.6 Conclusions

ICT use for learning at school has exploded in the last decades, and even further during the last years as a consequence of the COVID-19 pandemic. However, there is a lack of causal evidence, from sources that allow for comparability across different contexts and countries, on which is the impact of ICT use at school on student achievement. Previous analyses providing evidence comparable across countries from large-scale international surveys (in particular, those using data from PISA) did not incorporate research designs which allow to address endogeneity issues in a convincingly manner to identify the causal impact of ICT use on student achievement.

This paper estimates the impact of both ICT use and intensity of ICT use for learning at school on student achievement, in a cross-country analysis based on data from the most well-known international survey on educational processes and outcomes: PISA. To do so, this paper exploits within-student variability in both achievement and ICT use across two different subjects available in PISA: math and science. The main results obtained are twofold. First, the impact of both ICT use and intensity of ICT use for learning at school on students' outcomes in PISA varies depending on the country. Second, for a vast majority of countries, the impact of both ICT use and intensity of ICT use for learning at school on students' outcomes in PISA is non-significant. We carry out specific tests on the conditions needed to interpret these results as the causal impact of ICT use on student achievement. For those countries in which a causal interpretation of the relationship between ICT use for learning at school and students' outcomes in PISA is backed by the results of the tests, both results explained above are corroborated: the impact of ICT use for learning at school is heterogeneous and country-specific, and it is non-significant for a majority of countries.

The first of these results obtained in this paper connects with previous studies which highlight that the impact of ICT use on student achievement is context-specific: it depends on the country context (Vargas-Montoya et al., 2023), and it may also depend on the type of uses of ICT (Biagi and Loi, 2012; Comi et al., 2017; Falck et al., 2018). As regards the second result highlighted above, this paper diverges from the bulk of studies on the relationship between ICT use and students' outcomes based on PISA data which do not account for students' unobserved characteristics. Those studies tend to find a negative relationship between ICT use and PISA outcomes. In contrast, our result agrees with those from the majority of previous experimental and quasi-experimental studies on this issue, which obtained a non-significant impact of ICT use at school on student achievement (Angrist and Lavy, 2002; Rouse and Krueger, 2004; Goolsbee and Gurvan, 2006; Craig et al., 2013; Cristia et al., 2017). With respect to this literature, the most important contribution of our paper is that it provides evidence of the impact of ICT use on student achievement from a large-scale international survey (PISA), which allows that its results are representative at the student population level for the countries under analysis, and comparable with future studies carried out on this issue (as it uses standard measures of both student achievement and ICT use available in PISA).

The implications of these results are also twofold. First, authorities and policy-makers in education should avoid too simplistic optimistic approaches to the introduction of ICT at school, under an implicit (or explicit) assumption that increasing the amount of ICT available or used in learning processes will automatically led to an increase in student achievement. Empirical evidence shows that, in most cases, the effects of such increase of ICT use at school will have non-significant effects on student achievement. Second, both researchers and policy-makers should make further efforts to improve the understanding of in which contexts, under which circumstances and for which type of uses of ICT, an increase in ICT use at school will indeed have a positive impact on student achievement, as found in studies carried out in specific settings such as Banerjee et al. (2007), Barrow et al. (2009) and Bartelet et al. (2016).

General conclusions of the thesis

The purpose of having a world where quality of life reaches those who have been marginalized from the unprecedented prosperity of the last decades has brought together efforts by funders, multilateral organizations, non-profit organizations and other previous stakeholders in recent years. In this scenario, in which education plays a crucial role, different programs and interventions aimed at improving the educational outcomes of boys and girls have been implemented in different parts of the world, each of them focused on factors that are considered relevant in terms of promoting access and quality of education. Although the interventions and programs are designed and implemented by governments and organizations that have the best intentions of impacting educational outcomes and in general the well-being of the populations where they are implemented, it is necessary to use rigorous approaches that allow identifying what works and what does not. Based on administrative data and information collected at scale (i.e., PISA 2018), this thesis uses impact evaluation techniques to study, from an economic perspective, the expected and unexpected effects of public policies on educational outcomes in four different domains.

Main results

Chapter 2 presents the first experimental evaluation of the relative effectiveness of two approaches to improve school management in a developing country. In this experiment, a group of public primary schools in seven Mexican states were randomly assigned to receive training either directly from professional trainers or a cascade model of "training of trainers." The results reveal that, compared to indirect training, direct training improved the managerial capacity of school directors, although it failed to significantly improve learning outcomes. The chapter explores the potential channels that explain not finding effects on learning outcomes, and reasons associated with changes in the educational actions and assets of parents at home and the appropriation of the program by the participating directors are ruled out. The results considering the existing literature reveal that an impact on student learning required a major change in the managerial skills of the directors, which could potentially be limited by the low autonomy that schools have in Mexico. Given the cost of the intervention (~470 USD per school), and the lack of learning outcomes, resources could be redirected to interventions that focus on improving pedagogy (for example, teaching at the appropriate level, teaching content, and pedagogical content) and improve teacher accountability.

Chapter 3 focuses on a fundamental factor in the quality of education: the teachers. Using administrative data from nearly 25,000 new teachers in Mexico between 2012 and 2017, this chapter studies the effects of a reform that introduced the need for competitive tests for admission to the teaching service. The results show that the implementation of the reform improved between 3% and 6% the cognitive skills of the new teachers. This improvement was mainly driven by changes in the lower part of the skills distribution of newly hired teachers: for those new teachers who come from the 0.1 quantile of the skills distribution, an improvement of between 25% and 30% was observed, in contrast, the improvement for those coming from the 0.9 quantile was 1.5%. We propose two main channels to explain this result. First, the reform reduced the prevalence of discretionary hiring, from around 35% in the years prior to the reform to less than 13% after 3 years of its implementation. Second, the reform improved recruitment selection efficiency, making cognitive skills a more important determinant of recruitment outcomes for those using the meritocratic access system, while prior to the reform a potential teacher approaching the meritocratic system and coming from the highest part of the distribution had less than 20% probability of being hired, this same group happened to have close to 50% probability of being hired after the reform was implemented.

The reform decreased, but did not eliminate, the use of discretionary hiring. In addition, the cognitive skills of teachers who entered through the discretionary system deteriorated significantly. The incomplete scope of these reforms speaks to the challenges of implementing civil service reforms in contexts where institutions may be weak and incentives for non-compliance are strong. The results presented in this chapter show how a rules-based civil service system can improve the skills profile of incoming civil servants. However, to achieve lasting effects it must overcome underlying political obstacles to implementation, particularly in the context of a developing country.

Chapter 4 focuses on a public policy that, although it was not directly focused on educational outcomes, ended up having an influence on the educational trajectories of the most vulnerable children and youth living in the Colombian municipalities historically most affected by the armed conflict: the ceasefire with the FARC-EP. The results reveal that the de-escalation of armed confrontations that take place within the framework of ceasefires in armed conflicts is not necessarily accompanied by improvements in the education of the population strongly affected by the conflict. In fact, in this case, the opposite happened. Contrary to intuition, the results obtained reveal that the de-escalation of the armed conflict in Colombia negatively impacted the school attendance of students who live in the municipalities historically affected by the actions of the FARC-EP.

The main results reveal that the risk of desertion of children and young people from the most vulnerable households increased by 13 percent compared to the control municipalities. These results are concentrated among children and youth living in rural areas of the municipalities, with a 24 percent higher risk of dropping out compared to control municipalities, while for children and youth living in urban areas there is no such an effect. As the main mechanism to explain this result, we identified the rapid increase in illicit crops, which coincided with the de-escalation of the conflict in the municipalities historically most affected by violence in Colombia. For rural areas, where illicit crops are located and where the most vulnerable populations also live, the effect of the de-escalation of the conflict on the risk of desertion is fully explained by the presence of illicit crops. Conversely, for students living in urban areas, reduced conflict is associated with reduced risk of dropout.

Chapter 5 estimates the impact of both ICT use and the intensity of ICT use for learning at school on student performance. To do so, this chapter exploits the within-student variability in both performance and use of ICT between two different subjects available in PISA 2018: mathematics and science. The main results obtained are two. First, the impact of both ICT use and the intensity of ICT use for learning at school on student performance in PISA varies by country. Second, for a large majority of countries, it is found that the impact of both the use of ICT and the intensity of the use of ICT for learning at school on student results in PISA is not significant. For those countries where a causal interpretation of the relationship between the use of ICT for learning at school and student performance in PISA is supported by test results, both previously mentioned results are corroborated.

In terms of policy recommendations, the results lead to two messages. In the first place, the authorities and those responsible for formulating education policies must take into account that the use of ICT does not necessarily lead to an improvement of learning of the students who use them. Empirical evidence shows that, in most cases, the effects of such increased use of ICT at school will have insignificant effects on student performance. Second, both researchers and policy makers should do more to improve understanding of in what contexts, under what circumstances, and for what kinds of ICT uses, increased use of ICTs at school will have a positive impact on student achievement.

The expected and unexpected impacts explored through the different chapters of this thesis confirm that programs and interventions that, if considering their design and implementation, are expected to have a guaranteed impact, do not necessarily have it once they are evaluated through rigorous quantitative techniques. Results obtained within the framework of this thesis show, for example, that although the literature has shown the importance of school directors' skills for student learning, investing large amounts of resources in programs addressed at improving skills of schools' principals may not have the expected impact on learning if school autonomy is limited and the possibility of large changes in the allocation of resources within schools is beyond the control of principals. Another example is that, although it is widely accepted that peace settings are conducive to the development of skills through successful educational trajectories, long-term conflicts leave social and economic practices in the territory that, even in peace settings, can keep hindering the educational development of children and young people.

Future research agenda

The possibility of accessing large amounts of data and the appearance of impact evaluation techniques have made possible a rigorous approach to questions such as those addressed in this thesis; however, it is important to note that the different chapters of this thesis faced challenges in terms of access to sources of information, which affect some aspects of the analysis of the question posed or the possibility of exploring further related questions.

In Chapter 2, a significant limitation is related to the lack of information that detailed the work of the teachers in the schools under evaluation. Availability of that information would have made it possible to explore whether teachers had any changes in their teaching practices, use of time, or curricular work because of the new managerial skills of the principals.

In the case of Chapter 3, having access to information on the performance of Mexican students in standardized tests would have allowed to explore whether the teachers hired through the meritocratic mechanism had a positive impact on learning outcomes of their students, in contrast to teachers hired via the discretionary mechanism.

For Chapter 4, obtaining data on the labor markets in the different municipalities of the country would have helped to understand the flow of children to activities in agriculture as a proxy for work in coca leaf plantations.

Finally, Chapter 5 would have benefited if information that allowed pairing teachers and students was available in PISA 2018. That information would have served to understand if any teacher-specific pedagogical practice or scenario of ICT use at school was a significant determinant of the impact of ICT use on learning is.

The research agenda approached in this thesis aims to be further explored in the development of my career as a researcher. I am deeply interested in continuing to focus on understanding how programs, policies and interventions related to education affect the learning processes and outcomes of students from the most vulnerable contexts. Likewise, I am largely interested in studying the long-term effects of educational interventions carried out in early childhood, in order to understand how they affect employment trajectories, decisions on fertility and higher education achievements.

Conclusiones generales de la tesis

El propósito de tener un mundo donde la calidad de vida llegue a aquellos que han sido marginados del incremento de la prosperidad, que en las últimas décadas ha alcanzado un ritmo sin precedentes, ha aunado esfuerzos de financiadores, organizaciones multilaterales, organizaciones sin ánimo de lucro y otros actores involucrados en el tema. En este escenario, en el cual la educación desempeña un papel fundamental, los gobiernos han implementado diferentes programas e intervenciones orientadas a mejorar los resultados educativos de los niños, niñas y jóvenes en diferentes partes del mundo, cada una de ellas enfocadas a factores que se consideran relevantes en términos de promover el acceso y calidad de la educación. Las intervenciones y programas son diseñados e implementados por gobiernos y organizaciones que tienen las mejores intenciones de generar un impacto positivo en los resultados educativos y, en general, en el bienestar de las poblaciones donde se implementan. Sin embargo, es necesario utilizar aproximaciones rigurosas que permitan identificar qué funciona y qué no, con el fin de seguir avanzando hacia la selección y refinamiento de las políticas más efectivas. A partir de datos administrativos y grandes bases de datos de pruebas estandarizadas, esta tesis utiliza técnicas de evaluación de impacto para estudiar, desde una perspectiva económica, los efectos esperados y no esperados de políticas públicas en resultados educativos en cuatro ámbitos diferentes.

Resultados principales

El capítulo 2 presenta la primera evaluación experimental sobre la eficacia relativa de dos intervenciones para mejorar la gestión escolar en un país en desarrollo. En este experimento, un grupo de escuelas primarias públicas en siete estados mexicanos fue asignado aleatoriamente para recibir capacitación ya sea directamente de formadores profesionales o un modelo en cascada de "formación de formadores". Los resultados revelan que, en comparación con la capacitación indirecta, la capacitación directa mejoró la capacidad de gestión de los directores de escuela, pero no logró mejorar significativamente los resultados del aprendizaje. Este capítulo explora también los potenciales canales que explican no encontrar efectos en los resultados de aprendizaje y se descartan razones asociadas con cambios en las acciones y activos educativos de los padres en casa y la apropiación del programa por parte de los directores participantes. Los resultados, a la luz de la literatura existente, revelan que el impacto en aprendizajes de los estudiantes habría requerido un cambio mayor en las habilidades gerenciales de los directores, lo que potencialmente pudo estar limitado por la baja autonomía escolar que tienen las escuelas en México. Dado el costo de la intervención (~470 USD por escuela), y la ausencia de resultados sobre aprendizaje, los recursos podrían reorientarse a intervenciones que se centren en mejorar la pedagogía (por ejemplo, enseñanza en el nivel adecuado, contenido docente y contenido pedagógico) y mejorar la rendición de cuentas de los docentes.

El capítulo 3 se enfoca en un factor fundamental en la calidad de la educación: los docentes. Utilizando datos administrativos de cerca de 25.000 nuevos docentes en México entre el 2012 y 2017, este capítulo estudia los efectos de una reforma que implantó la necesidad de pruebas competitivas para el ingreso al servicio docente. Los resultados muestran que la implementación de la reforma mejoró entre el 3% y 6% las habilidades cognitivas de los nuevos docentes. Esta mejora fue impulsada principalmente por cambios en la parte inferior de la distribución de habilidades de los docentes recién contratados, para aquellos nuevos docentes situados en el percentil 0.1 de la distribución de habilidades se observó una mejora de entre el 25% y 30% en las habilidades cognitivas; en contraste, la mejora para los situados en el percentil 0.9 fue del 1.5%. El capítulo propone dos canales principales para explicar este resultado. En primer lugar, la reforma redujo la prevalencia de las contrataciones discrecionales, pasando de ser cerca del 35% en los años previos a la reforma a menos del 13% luego de 3 años de su implementación. En segundo lugar, la reforma mejoró la eficiencia de selección de la contratación, haciendo que las habilidades cognitivas sean un determinante más importante de los resultados de la contratación para aquellos que utilizan el sistema formal de acceso al servicio público: mientras que antes de la reforma un docente acercándose al sistema meritocrático y viniendo de la parte más alta de la distribución tenía menos del 20% de probabilidades de ser contratado, este mismo docente pasó a tener cerca del 50% de probabilidades de ser contratado después de la reforma.

La reforma disminuyó, pero no eliminó, el uso de la contratación discrecional; además, las habilidades cognitivas de los docentes que entraron por el sistema discrecional se deterioraron de manera significativa. El alcance incompleto de estas reformas refleja los desafíos de implementar reformas del servicio civil en contextos donde las instituciones pueden ser débiles y los incentivos para el incumplimiento, fuertes. Los resultados presentados en este capítulo muestran cómo un sistema de servicio civil basado en reglas puede mejorar el perfil de competencias de los funcionarios públicos entrantes; sin embargo, para lograr un efecto duradero, han de superarse las dificultades políticas que obstaculizan su implementación en el contexto de un país en desarrollo.

El capítulo 4 se enfoca en una política pública que, aunque no estuvo directamente focalizada en resultados educativos, terminó teniendo influencia sobre las trayectorias educativas de los niños, niñas y jóvenes que habitan en los municipios colombianos históricamente más afectados por el conflicto armado: el cese al fuego con las FARC-EP. Los resultados revelan que la desescalada de los enfrentamientos armados que suceden en el marco de un alto el fuego en conflictos armados no necesariamente va acompañado de mejoras en la educación de la población fuertemente afectada por el conflicto; de hecho, en este caso ocurrió lo contrario. Contrariamente a la intuición, los resultados obtenidos revelan que la desescalada del conflicto armado en Colombia ha impactado negativamente sobre la asistencia escolar de los estudiantes que viven en los municipios históricamente afectados por la acción de las FARC-EP.

Los principales resultados de este capítulo revelan que el riesgo de deserción de los niños y jóvenes de los hogares más vulnerables aumentó en un 13 por ciento en comparación con los municipios de control. Estos resultados se concentran entre los niños y jóvenes que viven en las zonas rurales de los municipios, con un riesgo de deserción un 24 por ciento mayor en comparación con los municipios de control, mientras que para los niños y jóvenes que viven en las zonas urbanas no se encontró tal efecto. Como principal mecanismo de este resultado, se identifica el rápido aumento de los cultivos ilícitos, que coincidió con la desescalada del conflicto en los municipios históricamente más afectados por la violencia en Colombia. Para las zonas rurales, donde se ubican cultivos ilícitos y donde también viven las poblaciones más vulnerables, el efecto de la desescalada del conflicto sobre el riesgo de deserción se explica plenamente por la presencia de cultivos ilícitos. Por el contrario, para los estudiantes que viven en áreas urbanas, la reducción del conflicto se asocia con una reducción del riesgo de abandono.

El capítulo 5 estima el impacto tanto del uso de las TIC como de la intensidad del uso de las TIC para el aprendizaje en la escuela sobre el rendimiento de los estudiantes. Para hacerlo, este capítulo explota la variabilidad en cada estudiante, tanto en el rendimiento como en el uso de las TIC, entre dos materias diferentes disponibles en PISA 2018: matemáticas y ciencias. Los principales resultados obtenidos son dos. En primer lugar, el impacto tanto del uso de las TIC como de la intensidad del uso de las TIC para el aprendizaje en la escuela sobre los resultados de los estudiantes en PISA varía según el país. En segundo lugar, para una gran mayoría de países, se obtiene que el impacto tanto del uso de las TIC como de la intensidad del uso de las TIC para el aprendizaje en la escuela sobre los resultados de los estudiantes en PISA no es significativo. Para aquellos países en los que una interpretación causal de la relación entre el uso de las TIC para el aprendizaje en la escuela y los resultados de los estudiantes en PISA está respaldada por los resultados de las pruebas, se corroboran ambos resultados mencionados previamente.

En términos de recomendaciones de política, estos resultados derivan en dos mensajes. En primer lugar, las autoridades y los responsables de la formulación de políticas educativas deben tener en cuenta que el uso de las TIC no deriva necesariamente en la mejora de los aprendizajes de los estudiantes que las utilizan. La evidencia empírica muestra que, en la mayoría de los casos, dichos aumentos del uso de las TIC en la escuela tendrán efectos no significativos en el rendimiento de los estudiantes. En segundo lugar, tanto los investigadores como los responsables de la formulación de políticas deberían esforzarse más para mejorar la comprensión de en qué contextos, bajo qué circunstancias y para qué tipo de usos de las TIC, un aumento en el uso de las TIC en la escuela tendrá un impacto positivo en el rendimiento de los estudiantes.

Los impactos esperados y no esperados explorados a través de los diferentes capítulos de esta tesis revalidan que programas e intervenciones que, por su diseño e implementación, se espera que tengan un impacto positivo, no necesariamente lo tienen una vez se evalúan a través de técnicas cuantitativas rigurosas. Los resultados obtenidos en el marco de esta tesis muestran, por ejemplo, que aunque la literatura ha mostrado la importancia que las habilidades de los directores de escuelas tienen en los aprendizajes de los estudiantes, invertir grandes cantidades de recursos en programas de habilidades de gestión para los directores de escuelas podría no tener el impacto esperado en aprendizajes si la autonomía escolar es limitada y la posibilidad de grandes cambios en la asignación de recursos dentro de las escuelas está fuera del control de los directores. Otro ejemplo se deriva de la observación de que, aunque es ampliamente aceptado que los escenarios de paz son propicios para la mejora de las trayectorias educativas, los conflictos de larga duración dejan en el territorio prácticas sociales y económicas que, también en escenarios de paz, pueden seguir siendo obstáculos para el desarrollo educativo de niños, niñas y jóvenes.

Agenda futura

La posibilidad de acceder a grandes cantidades de datos y el desarrollo de las técnicas de evaluación de impacto han hecho posible analizar de manera rigurosa preguntas como las que aborda esta tesis. Sin embargo, es importante señalar que los diferentes capítulos de la tesis se enfrentaron a obstáculos relativos al acceso a fuentes de información, que limitaron el análisis de algunos aspectos de la pregunta planteada o la posibilidad de explorar preguntas relacionadas.

En el caso del capítulo 2, tener la posibilidad de acceder a información que detallara el trabajo de los docentes en las escuelas analizadas hubiera permitido explorar si estos tuvieron algún cambio en sus prácticas docentes, uso del tiempo, o trabajo curricular como resultados de las nuevas habilidades gerenciales de los directores.

Para el caso del capítulo 3, tener la posibilidad de acceder al desempeño de los estudiantes mexicanos en las pruebas estandarizadas hubiera permitido explorar si los docentes que fueron contratados a través del mecanismo meritocrático tuvieron un impacto en los aprendizajes de sus estudiantes diferente al de los que entraron por la vía discrecional.

Para el capítulo 4, la disponibilidad de datos sobre los mercados laborales en los diferentes municipios del país podría hacer ayudado a entender el flujo de los niños y niñas a actividades en agricultura como un proxy del trabajo en las plantaciones de hoja de coca.

Finalmente, el capítulo 5 podría haberse beneficiado de la existencia de información en PISA 2018 que permitiera emparejar profesores y estudiantes, para analizar si alguna práctica pedagógica o escenario de uso de TIC en la escuela son determinantes respecto a los impactos que el uso de TIC pueda tener en los aprendizajes.

Espero, en el desarrollo de mi carrera como investigador, continuar profundizando en la agenda de investigación abordada en esta tesis. En particular, tengo gran interés en continuar analizando cómo los programas, políticas e intervenciones relacionadas con la educación inciden sobre los aprendizajes de los estudiantes en los contextos más vulnerables. Así mismo, tengo interés en estudiar los efectos a largo plazo de intervenciones educativas realizadas en los primeros años de la infancia, en particular, para analizar cómo inciden sobre las trayectorias laborales y sobre las decisiones en temas de fertilidad y educación superior.

References

Acemoglu, D. and Robinson J. (2013). Why Nations Fail. London, England: Profile Books.

Agüero, J. M. and Majid, M. F. (2014). War and the Destruction of Human Capital (No. 163). Retrieved from Households in Conflict Network website: https://ideas.repec.org/p/hic/wpaper/163.html.

Akresh, R. and de Walque, D. (2008). Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide (No. WPS4606; pp. 1–38). Retrieved from The World Bank website: http://documents.worldbank.org/curated/en/368931468144008432/Armed-conflict-and-schooling-evidence-from-the-1994-Rwandan-genocide.

Alderete, M.V.; Di Meglio, G. and Formichella, M.M. (2017): Acceso a las TIC y rendimiento educativo: ¿una relación potenciada por su uso? Un análisis para España. *Revista de Educación*, 377, 54-81.

Aleteia (2016,11,07). Escuelas vacías y niños “raspando” coca… ¿Hasta cuándo? Retrieved from https://es.aleteia.org/2016/07/11/escuelas-vacias-y-ninos-raspando-coca-hasta-cuando/

Álvarez, J., García-Moreno, V., & Patrinos, H. A. (2007). Institutional effects as determinants of learning outcomes: Exploring state variations in Mexico (Vol. 4286). World Bank Publications.

Angrist, J. and Lavy, V. (2002) New Evidence on Classroom Computers and Pupil Learning. *The Economic Journal*, 112(482), 735-765.

Angrist, J. D., and Kugler, A. D. (2008). Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia. *The Review of Economics and Statistics*, 90(2), 191–215.

Araujo, M. D., Heineck Guido., and Cruz-Aguayo, Y., (2020) Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador, IZA Discussion Papers 13830, Institute of Labor Economics (IZA).

Arias, M. A., Ibáñez, A.M. and Zambrano, A. (2019). Agricultural Production Amid Conflict: Separating the Effects of Conflict into Shocks and Uncertainty. *World Development*, Elsevier, vol. 119(C): 165–84.

Arjona, A. (2016). Rebelocracy: Social Order in the Colombian Civil War (Ser. Cambridge studies in comparative politics). Cambridge, United Kingdom: Cambridge University Press. Available at: https://repository.urosario.edu.co/handle/10336/20517.

Azevedo, J. P., Rogers, F. H., Ahlgren, S. E., Cloutier, M.H., Chakroun, B., Chang, Gwang-Chol., Mizunoya, S., Reuge, N. J., Brossard, M., Bergmann, J. L., (2021). The State of the Global Education Crisis: A Path to Recovery (English). Washington, D.C.: World Bank Group.

Bandara, A., Dehejia, R., and Lavie-Rouse, S. (2015). The impact of income and non-income shocks on child labor: evidence from a panel survey of Tanzania. *World development*, 67:218–237.

Bandiera, O. (2019) Alleviating poverty with experimental research: The 2019 Nobel laureates. https://cepr.org/voxeu/columns/alleviating-poverty-experimental-research-2019-nobel-laureates

Banerjee, A.V., Cole, S., Duflo, E. and Linden, L. (2007): Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 122(3), 1235-1264.

Banerjee, A., and Duflo, E. (2012). Poor Economics. New York, NY: PublicAffairs.

Banerjee, A., Duflo, E. and Kremer, M., 2016. The Influence of Randomized Controlled Trials on Development Economics Research and on Development Policy. Paper prepared for "The State of Economics, The State of the World"

Bardhan, P., (2002) Decentralization of Governance and Development, *Journal of Economic Perspectives*, 16 (4), 185–205.

Barrow, L.; Markman, L. and Rouse, C.E. (2009) Technology's Edge: The Educational Benefits of

Computer-Aided Instruction. *American Economic Journal: Economic Policy*, 1(1), 52-74.

Bartanen, B., Grissom, J. A., & Rogers, L. K. (2019) The impacts of principal turnover. *Educational Evaluation and Policy Analysis*, 41(3), 350-374. Retrieved from https:// doi.org/10.3102/0162373719855044 doi: 10.3102/0162373719855044

Bartelet, D., Ghysels, J., Groot, W., Haelermans, C. and Maasen van den Brink, H. (2016) The Differential Effect of Basic Mathematics Skills Homework via a Web-Based Intelligent Tutoring System Across Achievement Subgroups and Mathematics Domains: A Randomized Field Experiment. *Journal of Educational Psychology*, 108(1), 1-20.

Becker, S. O., I. Grosfeld, P. Grosjean, N. Voigtländer, and E. Zhuravskaya. (2020). Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers. *American Economic Review*, 110(5): 1430–63.

Bedoya Ospina, J. G., García Jaramillo, S., Rodríguez Orgales, C., Sánchez Ayala, L. M., and Sánchez Torres, F. J. (2019). Trayectorias de desarrollo de los jóvenes en zonas de posconflicto: diagnóstico y propuestas de política. Universidad de los Andes, Facultad de Economía, CEDE.

Bender, S., Bloom, N., Card, D., Van Reenen, J., & Wolter, S. (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics*, 36(S1), S371- S409. Retrieved from https://doi.org/10.1086/694107 doi: 10.1086/694107

Besley, T. J., Burgess, R., Khan, Adnan., and Xu, Guo., (2021). Bureaucracy and Development, Working Paper 29163, National Bureau of Economic Research.

Biagi, F. and Loi, M. (2013) Measuring ICT Use and Learning Outcomes: evidence from recent econometric studies. *European Journal of Education*, 48(1), 28-42.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4), 1351-1408. doi: 10.1162/qjec.2007.122.4.1351

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1), 1–51.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic Journal*, 125(584), 647–674.

Bloom, N., Lemos, R., Sadun, R., Scur, D., & Van Reenen, J. (2014). The new empirical economics of management. *Journal of the European Economic Association*, 12(4), 835– 876.

Bloom, N., Propper, C., Seiler, S., & Van Reenen, J. (2015). The impact of competition on management quality: Evidence from public hospitals. *Review of Economic Studies*, 82(2), 457-489.

Brassiolo, P., Estrada R., and Fajardo G., (2020). My (running) mate, the mayor: Political ties and access to public sector jobs in Ecuador, *Journal of Public Economics*, 191 (C).

Brassiolo, P., Estrada R., and Fajardo G., (2021). Bureaucratic turnover across levels of government, CAF Working Paper #2021.

Brassiolo, P., Estrada R., and Fajardo G., and Martinez-Correa J., (2021). Family Rules: Nepotism in the Mexican Judiciary, CAF Working Paper #2021/09

Brück, T., Di Maio, M., and Miaari, S. H. (2014). Learning The Hard Way: The Effect of Violent Conflict on Student Academic Achievement. *Journal of the European Economic Association*. https://doi.org/10.1093/jeea/jvy051.

Bruns, B., & Luque, J. (2014). Great teachers: How to raise student learning in Latin America and the Caribbean. World Bank Publications.

Brutti, Zelda and Fabio Sánchez Torres, (2021) Turning around teacher quality in Latin America: Renewed confidence and lessons from Colombia, *Economic Analysis and Policy*

Cabras, S. and Tena Horrillo, J.D. (2016). A Bayesian non-parametric modeling to estimate student response to ICT investment. *Journal of Applied Statistics*, 43(14), 2627-2642.

Camacho González, A., and Conover, E. (2009). Manipulation of Social Program Eligibility: Detection, Explanations and Consequences for Empirical Research. Universidad de los Andes. Facultad de Economía. CEDE.

Carrillo, B. (2020). Present Bias and Underinvestment in Education? Long-Run Effects of Childhood Exposure to Booms in Colombia. *Journal of Labor Economics* 38(4)

Cascio, E. U. and Narayan, A. (2015). Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change. Technical report. Cambridge, MA: National Bureau of Economic Research.

Chamarbagwala, R., and Morán, H. E. (2011). The Human Capital Consequences of Civil War: Evidence from Guatemala. *Journal of Development Economics*, 94(1): 41–61.

Chen, S., Loayza, N., and Reynal-Querol, M. (2007). The Aftermath of Civil War. Research Working Paper, No. 4190. Retrieved from http://hdl.handle.net/10986/7006

Chetty, R., Hendren, N., and Katz, L. F. (2015). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. Technical report. Cambridge, MA: National Bureau of Economic Research.

Chetty, R., John N. F., and Jonah E. R., (2014) Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood, *American Economic Review*, 104 (9), 2633–79.

Cingano, F. (2014), Trends in Income Inequality and its Impact on Economic Growth, OECD Social, Employment and Migration Working Papers, No. 163, OECD Publishing, Paris,

CNMH (Centro Nacional de Memoria Histórica). (2013). ¡Basta ya! Colombia, memorias de guerra y dignidad: informe general (Segunda edición corregida). Bogotá: Comisión Nacional de

Reparación y Reconciliación.

Colombian National Government (2016). Final Agreement to End the Armed Conflict and Build a Stable and Lasting Peace**.** http://especiales.presidencia.gov.co/Documents/20170620-dejacion-armas/acuerdos/acuerdo-final-ingles.pdf

Colonnelli, E., Mounu P., and Edoardo T., (2020). Patronage and Selection in Public Sector Organizations, *American Economic Review*, 110 (10), 3071–99.

Comi, S.L.; Argentin, G.; Gui, M.; Origo, F. and Pagani, L. (2017). Is it the way they use it? Teachers, ICT and student achievement. *Economics of Education Review*, 56, 24-39.

Corcoran, S. P., William N. E., and Robert M. S., (2004). Changing labor- market opportunities for women and the quality of teachers, 1957-2000, *American Economic Review*, 94 (2), 230–235.

Craig, S.D., Hu, X., Graesser, A.C., Bargagliotti, A.E., Sterbinsky, A., Cheney, K.R. and Okwumabua, T. (2013). The impact of a Technology-based Mathematics After-school Program using ALEKS on Student's Knowledge and Behaviors. *Computers & Education*, 68, 495-504.

Cristia, J.P.; Ibarrarán, P.; Cueto, S.; Santiago, A. and Severín, E. (2017): Technology and Child Development: Evidence from the One Laptop per Child Program. *American Economic Journal: Applied Economics*, 9(3), 295-320.

Dabalen, A. L., and Saumik, P. (2012). Estimating the Causal Effects of Conflict on Education in Côte d'Ivoire, World Bank Policy Research Working Paper 6077. Washington, DC: World Bank.

Dahis, Ricardo, Laura Schiavon, and Thiago Scot, (2020). Selecting Top Bureaucrats: Ad- mission Exams and Performance in Brazil, Working Paper, Available at SSRN

Dammert, A. C. (2008). Child labor and schooling response to changes in coca production in rural peru. *Journal of development Economics,* 86(1):164–180.

de Hoyos, R., Ganimian, A. J., & Holland, P. A. (2020). Great things come to those who wait: Experimental evidence on performance-management tools and training in public schools in Argentina.

de Hoyos, R., Ganimian, A. J., & Holland, P. A. (2019). Teaching with the test: Experimental evidence on diagnostic feedback and capacity building for public schools in Argentina. *The World Bank Economic Review*.

de Hoyos, R., García-Moreno, V., & Patrinos, H. A. (2017). The impact of an accountability intervention with diagnostic feedback: Evidence from Mexico. *Economics of Education Review*, 58, 123 - 140.

de Hoyos, Rafael, Ricardo Estrada, and María José Vargas, (2021). What do test scores really capture? Evidence from a large-scale student assessment in Mexico, *World Development*, 146, 105524.

De Witte, K. and Rogge, N. (2014). Does ICT matter for effectiveness and efficiency in mathematics education? *Computers & Education*, 75, 173-184.

Deaton, A. 2013. The Great Escape: health, wealth, and the origins of inequality. Princeton: Princeton University Press.

Di Maio, M., and Nandi, T. K. (2010). The Effect of the Israeli–Palestinian Conflict on Child Labor and School Attendance in the West Bank. *Journal of Development Economics* 100(2013): 107–116.

Dirección General de Planeación, Programación y Estadística Educativa. (2018). Sistema educativo de los Estados Unidos Mexicanos, principales cifras 2017-2018 (Tech. Rep.). Secretaría de Educación Pública.

Diwakar V. (2015). The Effect of Armed Conflict on Education: Evidence from Iraq. *Journal of Development Studies*, 51(12): 1702–18.

Dobbie, W., & Fryer, R. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, 5(4), 28-60.

Drew B. C., Mishra, A. and Brown, Annette., (2016) The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 8:1, 1-21.

Duncan, G (2013). La división del trabajo en el narcotráfico: mercancía, capital y geografía del Estado, in Giraldo, J. (ed.) 113–160. Economías criminales y poder político.

El Espectador. (2019, 06). Escuelas en ruinas y rodeadas de coca, así estudian en la cordillera de Nariño.

El Espectador. (2019, 10). En El Tandil, Tumaco, los profesores los paga la coca.

El Tiempo. (2015, 11). Los resultados de los CEASEs al fuego decretados por las Farc. Retrieved from https://www.eltiempo.com/archivo/documento/CMS-16384948

Elacqua, G., Iribarren, M. L., & Santos, H. (2018). Private schooling in Latin America: Trends and public policies (Tech. Rep.). Inter-American Development Bank. doi: http://dx.doi.org/10.18235/0001394

Erdogdu, F. and Erdogdu, E. (2015) The impact of access to ICT, student background and school/home environment on academic success of students in Turkey: An international comparative analysis. *Computers & Education*, 82, 26-49.

Estrada, Ricardo and Maria Lombardi, (2020) "Skills and selection into teaching: Evidence from Latin America," CAF Working Paper #2020/06

Estrada, Ricardo, "Rules versus discretion in public service: teacher hiring in Mexico," (2019) Journal of Labour Economics

Falck, O.; Mang, C. and Woessmann, L. (2018). Virtually No Effect? Different Uses of Classroom Computers and their Effect on Student Achievement. *Oxford Bulletin of Economics and Statistics*, 80(1), 1-38.

Fals-Borda, O., Guzmán-Campos, G., and Umaña-Luna, E. (2005). La violencia en Colombia. Bogotá: Taurus.

Fariña, P.; San Martín, E.; Preiss, D.D.; Claro, M. and Jara, I. (2015). Measuring the relation between computer use and reading literacy in the presence of endogeneity. *Computers & Education*, 80, 176-186.

Fernández-Gutiérrez, M.; Gimenez, G. and Calero, J. (2020). Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish Autonomous Communities. *Computers & Education*, 157.

Ferraro, S. (2018) Is information and communication technology satisfying educational needs at school? *Computers & Education*, 122, 194-204.

FLACSO. (2019). Políticas exitosas de desarrollo profesional docente en América Latina y el Caribe 2005-2016. Caracas: CAF. Retrieved from http://scioteca.caf.com/handle/123456789/1437

Fredriksson, P. and Ockert, B., (2008) The supply of skills to the teacher profession, *Scandinavian Journal of Economics*, 110 (2), 277–296.

Fryer, R. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics*, 129(3), 1355– 1407.

Fryer, R. (2017, May). Management and student achievement: Evidence from a randomized field experiment (Working Paper No. 23437). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w23437 doi: 10.3386/w23437

Gamazo, A. and Martínez-Abad, F. (2020) An exploration of factors linked to academic performance in PISA 2018 through data mining techniques. Frontiers in Psychology, 11, 575167.

García S., Rodríguez C., Sánchez F. and Bedoya J. G. (2015). La loteria de la cuna, Universidad de los Andes, Bogotá, Colombia.

Garcia, S. and Saavedra, J., (2022). Conditional Cash Transfers for Education. NBER Working Paper No. w29758, Available at SSRN: https://ssrn.com/abstract=4034209

Gates, S., Hegre, H., Nygård, H., Mokleiv & Strand, H., (2012). Development Consequences of Armed Conflict, *World Development*, Elsevier, vol. 40(9), pages 1713-1722.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. and Vermeersch, C. M. J. (2016). Impact Evaluation in Practice, Second Edition. © Washington, DC: Inter-American Development Bank and World Bank.

Glewwe, P., & Muralidharan, K. (2016). Chapter 10 - improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In S. M. Eric A. Hanushek & L. Woessmann (Eds.), (Vol. 5, p. 653 - 743). Elsevier.

Goolsbee, A. and Guryan, J. (2006). The Impact of Internet Subsidies on Public Schools. *The Review of Economics and Statistics*, 88(2), 336-347.

Grindle, M. S., (2012). Jobs for the Boys Patronage and the State in Comparative Perspective, Cambridge, MA: Harvard University Press

Gronqvist, E. and Jonas V., (2016)- One size fits all? The effects of teachers' cognitive and social abilities on student achievement, *Labour Economics*, 42, 138 – 150.

Guariso, A and Verpoorten, M. (2019). Armed Conflict and Schooling in Rwanda: Digging Deeper. *Peace Economics, Peace Science, and Public Policy*, De Gruyter, 25(1): 1–40.

Hanushek, E. A., Piopiunik M., and Wiederhold S., (2019) The value of smarter teachers international evidence on teacher cognitive skills and student performance, *Journal of Human Resources*, 54 (4), 857–899.

Haveman, R., & Smeeding, T. (2006). The Role of Higher Education in Social Mobility. *The Future of Children*, 16(2), 125–150. http://www.jstor.org/stable/3844794

Hertz, T., T, Jayasundera, P. Piraino, S. Selcuk, N. Smith & A. Verashchagina (2007). The inheritance of educational inequality: International Comparisons and Fifty-Year Trends, *The B.E. Journal of Economic Analysis and Policy*, Vol 7, Issue 2 (Advances), Article 10.

Hoeffler, A., and Reynal-Querol, M. (2003). Measuring the Costs of Conflict. Sambanis (2003). Breaking the Conflict Trap: Civil War and Development Policy. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.2979andrep=rep1andtype=pdf.

Hopkins, D., Ahtaridou, E., Matthews, P., Posner, C., & Figueroa, D. (2007). Reflections on the performance of the Mexican education system. OCDE. Directorate for Education, disponible en www. sep. gob. mx/work/models/sep1/Resource/93128/5/Mex PISA- OCDE2006.pdf

Hu, X.; Gong, Y.; Lai, C. and Leung, F.K.S. (2018). The relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A multilevel analysis. *Computers & Education*, 125, 1-13.

Ibañez, A. M., and Moya A. (2010). Do Conflicts Create Poverty Traps? Asset Losses and Recovery for Displaced Households in Colombia. NBER Chapters, in: The Economics of Crime: Lessons for and from Latin America, pp. 137–172. Cambridge, MA: National Bureau of Economic Research.

Ibáñez, A. M., and Velásquez, A. (2008). El impacto del desplazamiento forzoso en Colombia: condiciones socioeconómicas de la población desplazada, vinculación a los mercados laborales y políticas públicas. Políticas Sociales 145. Santiago: Economic Commission for Latin America and the Caribbean.

INEGI. (2018). Marco geoestadístico de Mexico. Retrieved 06/01/2018, from https://www.inegi.org.mx/temas/mg/default.html#

Instituto Nacional para la Evaluación de la Educación. (2018). Resultados de PLANEA (Tech. Rep.). Retrieved from https://www.inee.edu.mx/evaluaciones/planea/ resultados-planea/

International Labour Organization (2018) Guide on Measuring Decent Jobs for Youth - Monitoring, evaluation and learning in labour market programmes. Note 3: Establishing a monitoring system

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, 127(605), F236-F265.

Islas, P. M., Calef A. K., and Aparicio C., (2013). Mexico's Education Reform: A Multi-dimensional Analysis, Cham: Springer International Publishing, *Journal of Labor Economics*, 37 (2), 545–579.

Junguito Bonnet, R., Perfetti del Corral, J. J., and Delgado Barrera, M. (2017). Acuerdo de Paz: reforma rural, cultivos ilícitos, comunidades y costo fiscal. Bogotá: Fedesarrollo Centro de Investigación Económica y Social.

Justino, P. (2016). Supply and Demand Restrictions to Education in Conflict-affected Countries: New Research and Future Agendas. *International Journal of Educational Development*, 47: 76–85. https://doi.org/10.1016/j.ijedudev.2016.01.002

Kalyvas, S. N. (2015). How Civil Wars Help Explain Organized Crime - And How They Do Not. *Journal of Conflict Resolution*, 59(8): 1517–40.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130), 297–300.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *Economic Journal*, 125(588), 397-424.

Le Billon, P., Sherman, J., and Hartwell M. (2002). Controlling Resource Flows to Civil Wars: A Review and Analysis of Current Policies and Legal Instruments. International Peace Academy: Economic Agendas in Civil Wars: Policies and Practices for Regulating Resource Flows to Armed Conflict. New York, NY: International Peace Institute.

Lee, J. and Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, 65, 50–64.

Lemos, R., & Scur, D. (2016). Developing management: An expanded evaluation tool for developing countries. RISE Working Paper, 16(007).

Lemos, R., Muralidharan, K., & Scur, D. (2021). Personnel management and school productivity: Evidence from India (Working Paper No. 28336). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w28336 doi:10.3386/w28336

Leon, G. (2012). Civil Conflict and Human Capital Accumulation the Long-term Effects of Political Violence in Perú. *The Journal of Human Resources*, 47: 991–1022. 10.1353/jhr.2012.0036

Leuven, E.; Lindahl, M.; Oosterbeek, H. and Webbink, D. (2007) The Effect of Extra Funding for Disadvantaged Pupils on Achievement. *The Review of Economics and Statistics*, 89(4), 721-736.

Lochner, L and Moretti, E. (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review*, 94(1): 155–89.

Maldonado, J. H., Moreno-Sánchez, R. d., Espinoza, S., Bruner, A., Garzón, N., & Myers, J. (2018). Peace is Much More than Doves: The Economic Benefits of Bird-based Tourism as a Result of the Peace Treaty in Colombia. *World Development*, 106(C): 78–86.

McKenzie, D., & Ozier, O. (2019). Why ex-post power using estimated effect sizes is bad, but an ex-post mde is not. World Bank Development Impact Blog.

Meernik, J., DeMeritt, J., and Uribe-López, M. (2019). Introduction. In J. Meernik, J. DeMeritt, and M. Uribe-López (Eds.). As War Ends: What Colombia Can Tell Us About the Sustainability of Peace and Transitional Justice (pp.1–21). Cambridge: Cambridge University Press. doi:10.1017/9781108614856.001.

Mejía, D., and Restrepo P. (2013). Bushes and Bullets: Illegal Cocaine Markets and Violence in Colombia. Documentos CEDE 011934, Universidad de los Andes, Bogotá, Colombia.

Mejia, D., Prem, M., and Vargas, J. F. (2019). The Rise and Persistence of Illegal crops: Evidence from a Naive Policy Announcement. Serie Documentos CEDE: 2019–39. Universidad de los Andes, Bogotá, Colombia.

Mediavilla, M. and Escardíbul, J.O. (2015). ¿Son las TIC un factor clave en la adquisición de competencias? Un análisis con evaluaciones por ordenador. *Hacienda Pública Española*, 212, 67-96.

Merrouche, O. (2006). The Human Capital Cost of Landmine Contamination in Cambodia, HiCN Working Paper 25. The Households in Conflict Network, www.hicn.org.

Metzler, J. and Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics*, 99(2), 486-496.

Miguel, E., and Roland, G. (2005). The Long-run Impact of Bombing Vietnam. *Journal of Development Economics*. 96(1): 1–15.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, 36, 60–72.

Minoiu, C., and Shemyakina, O. N. (2014). Armed Conflict, Household Victimization, and Child Health in Côte d'Ivoire. *Journal of Development Economics*, 108(C): 237–55.

Mitchell, I. (2021). Measuring Development Cooperation and the Quality of Aid. In: The Palgrave Handbook of Development Cooperation for Achieving the 2030 Agenda. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-57938-8_11

Mookherjee, D., (2015). Political Decentralization, *Annual Review of Economics*, 7 (1), 231–249.

Moreira, D., and Pérez S., (2021). Civil Service Reform and Organizational Practices: Evidence from the Pendleton Act, Working Paper 28665, National Bureau of Economic Research

Moya, A, (2018). Violence, Psychological Trauma, And Risk Attitudes: Evidence From Victims Of Violence in Colombia. *Journal of Development Economics*, 131(C): 15–27.

Munoz, P., and Mounu P., (2021). Managers' Productivity and Recruitment in the Public Sector: The Case of School Principals, Technical Report

Muralidharan, K., & Singh, A. (2020). Improving public sector management at scale? experimental evidence on school governance India (Working Paper No. 28129). National Bureau of Economic Research. Retrieved from http://www.nber.org/ papers/w28129 doi: 10.3386/w28129

Namen, O., Prem, M, and Vargas, J. F. (2021). The Human Capital Peace Dividend Available at SSRN: https://ssrn.com/abstract=3424863 or http://dx.doi.org/10.2139/ssrn.3424863.

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D. G., Ramasubbaiah, R. G. N., and Thewissen, S. (2018). Fair Progress?: Economic Mobility Across Generations Around the World. Equity and Development. © Washington, DC: World Bank

Neilson, C., Gallegos S., and Calle F., (2019). Screening and recruiting talent at teacher colleges using pre-college academic achievement, Working Paper

Newman, J. L., and Gertler, P. J. (1994). Family Productivity, Labor Supply, and Welfare in a Low-Income Country. *The Journal of Human Resources*, 29(4): 989–1026.

Nickell, S., and Glenda Q., (2002) The consequences of the decline in public sector pay in Britain: a little bit of evidence, *The Economic Journal*, 112 (477), F107–F118.

OECD (2015) Students, Computers and Learning. Making the Connection. Paris: OECD Publishing.

OECD (2020): PISA 2018 Technical Report. https://doi.org/10.1787/baa65ceb-en

OECD (2023), Equity and Inclusion in Education: Finding Strength through Diversity, OECD Publishing, Paris

OECD (2016). PISA 2015 results (volume II): Policies and practices for successful schools.

Ornaghi, Arianna, (2019) "Civil service reforms: Evidence from US police departments,"

Oyelere, R. T. and Wharton K. (2013). The Impact of Conflict on Education Attainment and Enrollment in Colombia: Lessons from Recent IDPs. HiCN Working Papers 141, Households in Conflict Network.

Parlow, A. (2012). Armed Conflict and Children's Health. Exploring New Directions: The Case of Kashmir, MPRA Paper No. 38033.

Petko, D.; Cantieni, A. and Prasse, D. (2017). Perceived Quality of Educational Technology Matters: A Secondary Analysis of Students' ICT Use, ICT-Related Attitudes, and PISA 2012 Test Scores. Jo*urnal of Educational Computing Research*, 54(8), 1070-1091.

Pizarro, E. (2011). Las FARC (1949–2011): de guerrilla campesina a máquina de guerra. Bogotá: Norma.

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2018). Teacher professional development around the world: The gap between evidence and practice (Tech. Rep.). The World Bank.

Prem, M., Rivera, A., Romero, D., and Vargas, J. F. (2019): Civilian Selective Targeting: The Unintended Consequences of Partial Peace. Available at SSRN 3203065.

Prem, M., Saavedra, S., and Vargas, J. F. (2020). End-of-conflict Deforestation: Evidence from Colombia's Peace Agreement. *World Development*, 129.

Rauch, J., (1995). Bureaucracy, Infrastructure, and Economic Growth: Evidence from U.S. Cities During the Progressive Era, *The American Economic Review*, 85 (4), 968-79.

Ravallion, M., 2016. The Economics of Poverty: History, Measurement, and Policy, OUP Catalogue, Oxford University Press,

Rivkin, S.G. and Schiman, J.C. (2015). Instruction Time, Classroom Quality, and Academic Achievement. *Economic Journal*, 125(588), 425-448.

Rivkin, Steven G, Eric A Hanushek, and John F Kain, (2005). Teachers, schools, and academic achievement, *Econometrica*, 73 (2), 417–458.

Rockoff, J. E., Jacob B. A., Kane T. J., and Staiger D. O., (2011). Can you recognize an effective teacher when you recruit one?, *Education finance and Policy*, 6 (1), 43–74.

Rodríguez, C., and Sánchez, F. (2012). Armed Conflict Exposure, Human Capital Investments, And Child Labor: Evidence from Colombia. Defence & Peace Economics, 23(2), 161–184. https://doi.org/10.1080/10242694.2011.597239.

Rodríguez, C., and Sánchez, F. J. (2010). Books and guns: the quality of schools in conflict zones. Universidad de los Andes, Facultad de Economía, CEDE

Ross, M. L. (2004). What do we know about natural resources and civil war? *Journal of peace research*, 41(3), 337–356.

Rouse, C.E. and Krueger, A.B. (2004). Putting computerized instruction to the test: a randomized evaluation of a “scientifically based” reading program. *Economics of Education Review*, 23, 323-338.

Skryabin, M.; Zhang, J.; Liu, L. and Zhang, D. (2015). How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education*, 85, 49-58.

Saavedra, S., Ladino, J. F., and Wiesner, D. (2019). One Step Ahead of the Law: The Net Effect of Anticipation and Implementation of Colombia’s Illegal Crops Substitution Program. Universidad del Rosario. Facultad de Economía, Bogotá, Colombia.

Sánchez, F., and Díaz A. M. (2007). Geografía de los Cultivos Ilícitos y Conflicto Armado en Colombia. In Sanchez, F. (ed.). Las Cuentas de la Violencia. Ensayos Económicos Sobre el Crimen y el Conflicto en Colombia, pp. 131–205. Bogotá, Colombia: Norma.

Sánchez, G., and Meertens, D. (2011). Bandoleros, gamonales y campesinos. El caso de la violencia en Colombia. New York: Penguin Random House.

Santiago, P., McGregor, I., Nusche, D., Ravela, P., & Toledo, D. (2012). Oecd reviews of evaluation and assessment in education: Mexico 2012. OECD Publishing. doi: http:// dx.doi.org/10.1787/9789264172647-en

Santibanez, Lucrecia, (2008). Reforma Educativa: El papel del SNTE, *Revista Mexicana de Investigación Educativa*, 13 (037), 419–443.

Santos, R. J. (2018). Blessing and Curse. The Gold Boom and Local Development in Colombia. *World Development*, 106(C): 337–55.

Secretaría de Educación Pública, & Banco Internacional de Reconstrucción y Fomento. (2015). Evaluación de impacto del ejercicio y desarrollo de la autonomía de gestión escolar y estrategia de intervención controlada (Tech. Rep.). Retrieved from http://escuelaalcentro.com/wp-content/uploads/2018/ 04/Documento-base-Evaluaci%C3%B3n-de-Impacto.pdf

Shah, M., and Steinberg, B. M. (2017). Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks On Human Capital. *Journal of Political Economy*, 125(2): 527–61.

Shakya, A. (2010). Experiences of Children in Armed Conflict in Nepal. *Children and Youth Services Review* 33(2011): 557–63.

Shemyakina, O. (2011). The Effect of Armed Conflict On Accumulation Of Schooling: Results from Tajikistan. *Journal of Development Economics*, 95(2): 186–200. Available at: https://doi.org/10.1016/j.jdeveco.2010.05.002

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., Jimenez, E. (2016). The impact of education programmes on learning and school participation in low-and middle-income countries.

Spiezia, V. (2010). Does Computer Use Increase Educational Achievements? Student-level Evidence from PISA. *OECD Journal: Economic Studies*, 1-22.

Stallings, J. (1977). Learning to look: a handbook on classroom observation and teaching models. Wadsworth Pub. Co. Retrieved from https://books.google.com.mx/books?id= QEglAQAAIAAJ

Stallings, J., & Molhlman, G. (1988). Classroom observation techniques. In J. Keeves (Ed.), Educational research, methodology and measurement: An international handbook. Elsevier Science & Technology Books.

Stampini, M. and Tornarolli, L., 2012. The Growth of Conditional Cash Transfers in Latin America and the Caribbean: Did They Go Too Far?, IZA Policy Papers 49, Institute for the Study of Labor (IZA).

Stewart, F., Huang, C., and Wang, M. (2011). Internal Wars in Developing Countries: An Empirical Overview of Economic and Social Consequences. In War and Underdevelopment: Volume 1: The Economic and Social Consequences of Conflict. Oxford Scholarship Online.

Sviatschi, M. M. (2016). Effects of Childhood Exposure to Illegal Labor Markets: Kindling Criminal Paths. 67.

Sviatschi, M. M. (2018). Making a Narco: Childhood Exposure to Illegal Labor Markets and Criminal Life Paths. Working Papers 2018–03, Princeton University Woodrow Wilson School of Public and International Affairs, Princeton, NJ.

Molina Millán T., Barham T., Macours K., Maluccio J. A., Stampini M. (2019). Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence, *The World Bank Research Observer*, Volume 34, Issue 1, Pages 119–159, https://doi.org/10.1093/wbro/lky005

The New York Times. (July 18, 2017). Los campesinos de la coca en Colombia temen por su subsistencia en tiempos de paz. Retrieved from https://www.nytimes.com/es/2017/07/18/espanol/america-latina/colombia-sustitucion-cultivos-paz-farc.html

Trucano, M. (2016). SABER-ICT Framework Paper for Policy Analysis: Documenting national educational technology policies around the world and their evolution over time. World Bank Education, Technology & Innovation: SABER-ICT Technical Paper Series (#01). Washington, DC: The World Bank.

Ujhelyi, G., (2014) Civil Service Rules and Policy Choices: Evidence from US State Governments, *American Economic Journal: Economic Policy,* 6 (2), 338–80.

United Nations (2015). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible

UNODC (2016) Colombia Monitoreo de territorios afectados por cultivos ilícitos 2014. Vienna: UNODC.

UNODC (United Nations Office on Drugs and Crime. (2013). Colombia Monitoreo de territorios afectados por cultivos ilícitos 2012. Vienna: UNODC.

UNODC and Fundación ideas para la paz. (2018) ¿Quiénes son las familias que viven en las zonas con? Caracterización de las familias beneficiarias del Programa Nacional Integral de Sustitución de Cultivos Ilícitos (PNIS). Fundación Ideas para la Paz. Bogotá: Colombia. Retrieved from http://ideaspaz.org/media/website/FIP_familiascoca_final.pdf

Valente, C. (2014). Education and Civil Conflict in Nepal. W*orld Bank Economic Review*, 28(2): 354–83. https://doi.org/10.1093/wber/lht014.

Vargas-Montoya, L.; Gimenez, G. and Fernández-Gutiérrez, M. (2023). ICT use for learning and students' outcomes: Does the country's development level matter? *Socio-Economic Planning Sciences*, DOI: 10.1016/j.seps.2023.101550.

Viennet, R. and B. Pont (2017). Education policy implementation: A literature review and proposed framework, OECD Education Working Papers, No. 162, OECD Publishing, Paris,

Villamizar, D. (2017). Las guerrillas en Colombia (Primera edición). Bogotá: Penguin Random House Grupo Editorial, S.A.S.

Voth, Joachim and Guo Xu, (2020) Discretion and Destruction: Promotions, Performance, and Patronage in the Royal Navy, Mimeo

White, Howard & Raitzer, David. (2018). Impact Evaluation of Development Interventions: A Practical Guide.

World Bank. (2007). What is school-based management? (Tech. Rep. No. 44922). Retrieved from http://documents.worldbank.org/curated/en/113901468140944134/What-is-school-based management

World Bank. (2017a). Primary completion rate, total (% of relevant age group) (data retrieved from World Development Indicators,

World Bank. (2017b). School enrollment, primary (% net) (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.PRM.NENR?locations=MX)

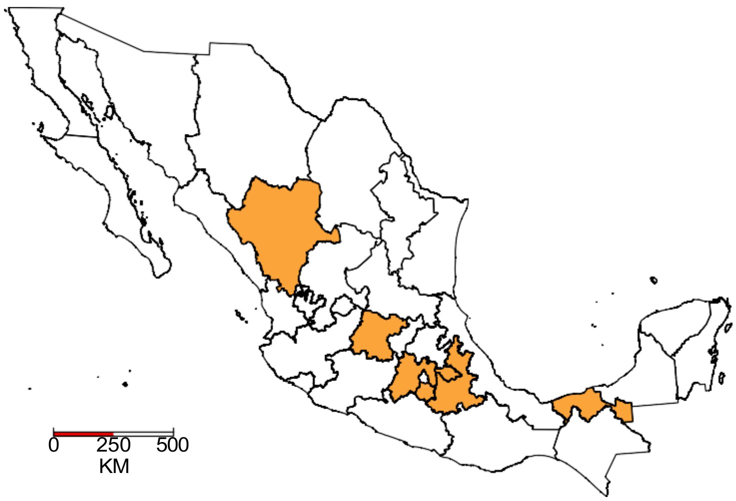
World Bank; Filmer, Deon P.; Rogers, F. Halsey; Al-Samarrai, Samer; Bendini, Maria Magdalena; Béteille, Tara; Evans, David; Kivine, Märt; Sabarwal, Shwetlena; Valerio, Alexandria; Abu-Jawdeh, Malek; Larson, Bradley Robert; Shrestha, Unika; Yuan, Fei. 2018. World Development Report (WDR) 2018: Learning to Realize Education's Promise Washington, D.C.: World Bank Group.

Xu, Guo, (2018) The Costs of Patronage: Evidence from the British Empire, *American Economic Review*, 108 (11), 3170–98.

Zhang, D. and Liu, L. (2016): How Does ICT Use Influence Students' Achievements in Math and Science Over Time? Evidence from PISA 2000 to 2012. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(9), 2431-2449.

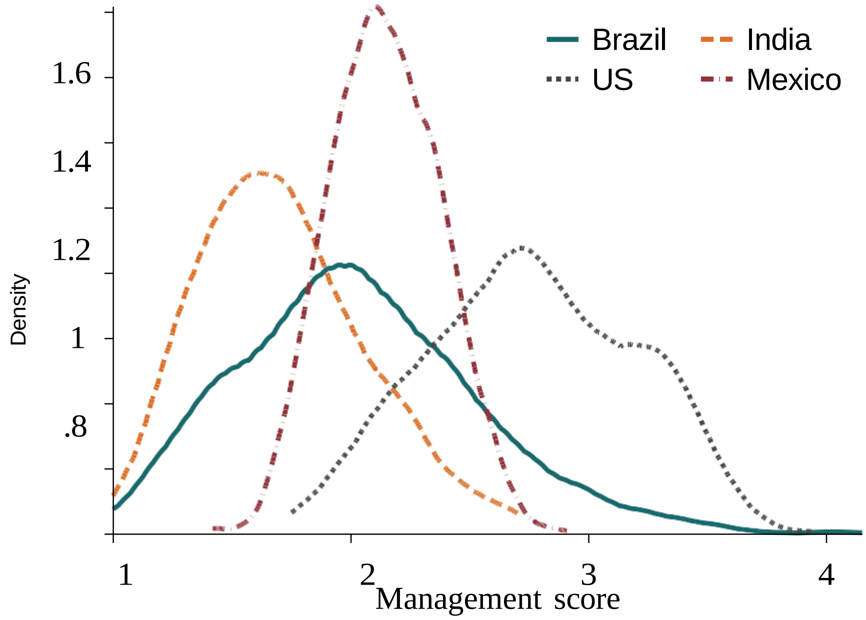
Appendix A: supplemental material chapter 2

Figure A.1: States participating in the impact evaluation



Note: Geographical information on the administrative areas of Mexico comes from INEGI (2018). Figure A.3 provide the distribution of schools within each state.

Figure A.2: Distribution of management practices in Brazil, India, Mexico, and the US



Note: Distribution of management practices from Brazil, India, and the US is based on the replication data of Bloom, Lemos, et al. (2015). The distribution of management practices from Mexico comes from our baseline data collected in 2015.

Table A.1: Balance between schools in the experimental sample and other schools

	(1)	(2)	(3)
	Mean		**Difference**
Variable	Participant	Non-participant	(1)-(2)
Students in math achievement L-IV (%)	8.08	10.95	-2.87***
	(11.70)	(17.79)	(0.36)
Students in math achievement L-I (%)	60.07	54.96	5.11***
	(22.01)	(28.04)	(0.67)
Students in language achievement L-IV (%)	2.99	4.93	-1.94***
	(5.28)	(10.68)	(0.17)
Students in language achievement L-I (%)	51.83	47.63	4.21***
	(20.38)	(28.02)	(0.62)
Marginalization	0.58	0.53	0.05***
	(0.49)	(0.50)	(0.01)
Urbanization	0.40	0.37	0.03**
	(0.49)	(0.48)	(0.01)
Number of students	279.55	196.79	82.75***
	(163.88)	(199.19)	(4.94)
Number of teachers	9.45	6.56	2.89***
	(4.31)	(5.72)	(0.13)
Student-teacher ratio	28.63	29.42	-0.79***
	(7.05)	(11.09)	(0.22)
Observations	1,194	20,611	21,805

This table presents the mean and standard error of the mean (in parentheses) for schools not in the experiment (Column 1) and those in the experiment (Column 2). Column 3 shows the mean difference between participant and non-participant schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to PLANEA exam results, which are scored from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is coded 1 for schools located in an urban area, and 0 otherwise. The number of students and teachers is taken from *Formato 911* from the year 2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Balance between schools with and without a DWMS audio file

	(1)	(2)	(3)
	Mean		**Difference**
Variable	No DWMS Audio	DWMS Audio	(1)-(2)
Direct training	0.50	0.52	-0.02
	(0.50)	(0.50)	(0.04)
Students in math achievement L-IV (%)	8.00	8.38	-0.61
	(11.78)	(11.45)	(0.78)
Students in math achievement L-I (%)	60.17	59.73	0.93
	(21.73)	(22.98)	(1.49)
Students in language achievement L-IV (%)	2.95	3.10	-0.25
	(5.27)	(5.33)	(0.37)
Students in language achievement L-I (%)	52.04	51.11	1.72
	(20.03)	(21.59)	(1.43)
Marginalization	0.58	0.55	0.02
	(0.49)	(0.50)	(0.03)
Urbanization	0.38	0.47	-0.06*
	(0.49)	(0.50)	(0.03)
Number of students	280.56	276.03	3.45
	(161.77)	(171.27)	(11.03)
Number of teachers	9.43	9.55	0.08
	(4.27)	(4.48)	(0.30)
Student-teacher ratio	28.90	27.69	0.41
	(6.98)	(7.22)	(0.43)
Observations	267	927	1,194

This table presents the mean and standard error of the mean (in parentheses) for schools without audio for the DWMS endline interview (Column 1) and schools with it (Column 2). Column 3 shows the mean difference between both types of schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to PLANEA exam results, which are scored from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the year 2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Differences in the likelihood of a DWMS audio file at the endline by school characteristics

	(1)	(2)	(3)
		DWMS Audio	
Direct training	-0.014		-0.045
	(0.024)		(0.11)
Students in math achievement L-IV (%)		-0.00076	0.00011
		(0.0012)	(0.0016)
Students in language achievement L-IV (%)		-0.00052	0.0022
		(0.0028)	(0.0047)
Marginalization		-0.0040	0.018
		(0.031)	(0.043)
Urbanization		-0.073**	-0.11**
		(0.033)	(0.045)
Student-teacher ratio		0.0031	0.0022
		(0.0021)	(0.0029)
Direct training × Students in math achievement L-IV (%)			-0.0017
			(0.0023)
Direct training × Students in language achievement L-IV (%)			-0.0030
			(0.0059)
Direct training × Marginalization			-0.044
			(0.055)
Direct training × Urbanization			0.064
			(0.059)
Direct training × Student-teacher ratio			0.0017
			(0.0037)
No. of obs.	1,194	1,193	1,193

This table presents the effect of different school characteristics on the likelihood the audio for the endline was usable. Achievement level (L) refers to PLANEA exam results, which are scored from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the year 2015. All regressions take into account the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Principal self-reported information from 2018 PLANEA-Contexto surveys

	Mean (SD)		Difference
	Train the trainer	Direct training	
Panel A: Courses or counseling on how to carry out school director duties			
Ever	0.74	0.85	0.11***
	(0.44)	(0.36)	(0.02)
Past 12 months	0.85	0.94	0.08***
	(0.35)	(0.24)	(0.02)
Panel B: Actions taken often or very often during the past 12 months			
Activities to improve learning	0.79	0.77	-0.02
	(0.41)	(0.42)	(0.03)
Classroom observations	0.71	0.72	0.01
	(0.45)	(0.45)	(0.03)
Help teachers improve pedagogical practices	0.79	0.81	0.03
	(0.41)	(0.39)	(0.02)
Provide parents with performance information	0.92	0.93	0.01
	(0.27)	(0.26)	(0.02)
Observations	565	545	1,110

This table presents the means and standard deviations (in parentheses) for "train the trainer" (Column 1) and "direct training" schools (Column 2). Column 3 presents the differences between groups, taking into account the randomization design (i.e., including strata fixed effects); standard errors (in parentheses) are clustered at the school level. Data come from PLANEA-Contexto questionnaires completed by school principals. Panel A includes self-reported information about courses or counseling on carrying out school director duties ever taken or in the past 12 months. Panel B indicates how often the principal engages in different practices (often and very often are coded 1, while sometimes and never are coded 0). "Activities to improve learning" is the teacher's self-reported frequency of taking any action to improve learning or the curriculum, including classroom observations, teacher evaluations, student evaluations, and acting as a tutor for teachers to improve pedagogical practices. "Classroom observations" is the self-reported frequency of such observations. "Help teachers improve pedagogical practice" is the self-reported frequency with which the principal helps the teacher improve their pedagogical practices. "Provide parents with performance information" is the self-reported frequency with which the teacher provides parents with school and student performance information. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Association between DWMS components and test scores at baseline across all schools

	(1)	(2)	(3)	(4)
	PLANEA 2015 scores			
Operations	-0.020			
	(0.025)			
Monitoring		0.0061		
		(0.024)		
Targets			0.00039	
			(0.026)	
People				-0.0087
				(0.028)
No. of obs.	20,049	20,049	20,049	20,049

This table presents the conditional correlation between DWMS components and student test scores at baseline across all schools. All regressions control for strata fixed effects, the number of pupils in the school, the pupil–teacher ratio, the marginalization index, and enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Effects on learning outcomes

	(1)	(2)	(3)	(4)
	PCA score			
Direct training	0.035	0.032	0.032	0.035
	(0.029)	(0.025)	(0.025)	(0.024)
No. of obs.	37,958	37,958	37,958	37,958
Lagged scores	No	Yes	Yes	Yes
Student controls	No	No	Yes	Yes
School controls	No	No	No	Yes

This table presents the treatment effects on learning outcomes (measured using PLANEA scores). The outcome is a composite index across subjects. All regressions take into account the randomization design (i.e., include strata fixed effects). "Lagged scores" indicates whether school average test scores from 2015 are included as controls. "Student controls" indicates whether age and gender are included as controls. "School controls" indi- cates whether the following controls are included: whether the school has a day shift, whether a primary school is intended to serve an indigenous population, the school's age, whether the school is located in an urban area, and the marginalization index of the school's municipality. Standard errors are clustered at the school level. * $p <$ 0.10, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Effect of DWMS on learning outcomes: Instrumental variable approach

	(1) Math	(2) Language	(3) Average	(4) PCA
DWMS	0.39	0.45	0.49	0.49
	(0.30)	(0.31)	(0.34)	(0.34)
No. of obs.	30,956	31,270	29,926	29,926

This table presents the effects of increasing the DWMS on learning outcomes (measured using PLANEA scores). We instrument the DWMS with the treatment assignment. The underlying assumption is that the DWMS completely captures any effect the treatment assignment might have on test scores. The first stage is presented in Panel A, Table 4. The outcomes are math test scores (Column 1), language test scores (Column 2), the average across subjects (Column 3), and a composite index across subjects (Column 4). All regressions take into ac- count the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Effects on other outcomes

	(1) Pass rate	(2) Repetition rate	(3) Enrollment
Direct training	0.0308	0.0251	13.65
	(0.253)	(0.114)	(8.734)
No. of obs.	1,186	1,185	1,192
Control mean	99.22	0.72	258.49

This table presents the treatment effects on the percentage of students who successfully complete their grade and can progress to the next one (pass rate in Column 1), the percentage of students that repeat a grade (Column 2), and the total number of students enrolled (Column 3). All outcomes refer to the 2017–2018 school year. All regressions take into account the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Balance between schools who answered the Stallings' implementation survey and those that did not

Variable	(1) Mean: Answered survey	(2) Mean: No survey answer	(3) Difference: (1)-(2)
Direct training	0.64	0.17	0.51***
	(0.48)	(0.38)	(0.03)
Students in math achievement L-IV (%)	8.41	7.30	1.85**
	(12.14)	(10.56)	(0.80)
Students in math achievement L-I (%)	60.07	60.06	-1.58
	(21.98)	(22.10)	(1.45)
Students in language achievement L-IV (%)	3.12	2.68	0.75**
	(5.69)	(4.13)	(0.36)
Students in language achievement L-I (%)	51.34	53.03	-3.32**
	(19.95)	(21.37)	(1.36)
Marginalization	0.59	0.56	0.01
	(0.49)	(0.50)	(0.03)
Urbanization	0.40	0.42	-0.02
	(0.49)	(0.49)	(0.03)
Number of students	282.89	271.45	17.25*
	(164.83)	(161.50)	(9.82)
Number of teachers	9.58	9.14	0.45*
	(4.35)	(4.20)	(0.26)
Student-teacher ratio	28.60	28.71	0.50
	(7.13)	(6.87)	(0.39)
Observations	845	349	1,194

This table presents the mean and standard error of the mean (in parentheses) for schools taking the Stallings implementation and use survey (Column 1) and schools not taking it (Column 2). Column 3 shows the mean difference between participant and non-participant schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to PLANEA exam scores, which range from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the year 2015. * $p < 0.10$, ** $p < 0.05$, $^{***}p < 0.01$

Table A.10: Likelihood of answering the Stallings' implementation survey by school characteristics

	(1)	(2)	(3)
	Answered Stallings' survey		
Direct training	0.39***		0.12
	(0.023)		(0.11)
Students in math achievement L-IV (%)		0.0019	0.0015
		(0.0014)	(0.0019)
Students in language achievement L-IV (%)		0.0032	-0.0038
		(0.0025)	(0.0055)
Marginalization		0.012	0.029
		(0.032)	(0.046)
Urbanization		-0.026	-0.017
		(0.032)	(0.048)
Student-teacher ratio		0.0030	-0.0030
		(0.0021)	(0.0029)
Direct training × Students in math achievement L-IV (%)			0.0017
			(0.0025)
Direct training × Students in language achievement L-IV (%)			0.0054
			(0.0059)
Direct training × Marginalization			-0.026
			(0.052)
Direct training × Urbanization			0.028
			(0.055)
Direct training × Student-teacher ratio			0.0084**
			(0.0034)
No. of obs.	1,194	1,193	1,193

This table presents the effect of different school characteristics on the likelihood of taking the Stallings' implementation survey by schools' characteristics. Achievement level (L) refers to PLANEA exam results, which are scored from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the year 2015. All regressions take into account the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Balance between schools who answered the SisAT's implementation survey and those that did not

Variable	(1) Mean: Answered survey	(2) Mean: No survey answer	(3) Difference: (1)-(2)
Direct training	0.53	0.41	0.13***
	(0.50)	(0.49)	(0.03)
Students in math achievement L-IV (%)	8.30	7.46	-0.35
	(11.67)	(11.81)	(0.85)
Students in math achievement L-I (%)	59.35	62.17	-0.39
	(22.04)	(21.83)	(1.51)
Students in language achievement L-IV (%)	3.08	2.73	0.17
	(5.66)	(3.96)	(0.28)
Students in language achievement L-I (%)	51.54	52.70	-0.03
	(20.40)	(20.33)	(1.38)
Marginalization	0.62	0.46	0.05*
	(0.49)	(0.50)	(0.03)
Urbanization	0.38	0.48	0.02
	(0.49)	(0.50)	(0.03)
Number of students	281.89	272.72	15.97
	(165.34)	(159.63)	(10.28)
Number of teachers	9.54	9.21	0.37
	(4.36)	(4.15)	(0.27)
Student-teacher ratio	28.60	28.74	0.63
	(7.08)	(6.98)	(0.40)
Observations	889	305	1,194

This table presents the mean and standard error of the mean (in parentheses) for schools taking the SisAT implementation and use survey (Column 1) and schools not taking it (Column 2). Column 3 shows the mean difference between participant and non-participant schools, as well as the standard error of the difference, clustered at the school level. Achievement level (L) refers to PLANEA exam scores, which range from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the year 2015. * $p < 0.10$, ** $p < 0.05$, $^{***}p < 0.01$

Table A.12: Differences in the likelihood of answering the SisAT's implementation survey by school characteristics

	(1)	(2)	(3)
	Answered SisAT's survey		
Direct training	0.089***		0.16
	(0.024)		(0.11)
Students in math achievement L-IV (%)		-0.00092	-0.0022
		(0.0014)	(0.0020)
Students in language achievement L-IV (%)		0.0024	0.0019
		(0.0023)	(0.0058)
Marginalization		0.061**	0.12***
		(0.031)	(0.045)
Urbanization		0.023	0.084*
		(0.033)	(0.048)
Student-teacher ratio		0.0029	0.0023
		(0.0020)	(0.0030)
Direct training × Students in math achievement L-IV (%)			0.0027
			(0.0026)
Direct training × Students in language achievement L-IV (%)			-0.00091
			(0.0062)
Direct training × Marginalization			-0.11**
			(0.055)
Direct training × Urbanization			-0.10*
			(0.058)
Direct training × Student-teacher ratio			0.00070
			(0.0036)
No. of obs.	1,194	1,193	1,193

This table presents the effect of different school characteristics on the likelihood of taking the SisAT's implementation survey by school characteristics. Achievement level (L) refers to PLANEA exam results, which are scored from L-I (lowest) to L-IV (highest). Marginalization is a variable coded 1 for areas that have a "high" or "very high" marginalization, and 0 otherwise according to CONAPO. Urbanization is a variable coded 1 for schools located in an urban area and 0 otherwise. The number of students and teachers is taken from *Formato 911* for the year 2015. All regressions take into account the randomization design (i.e., include strata fixed effects). Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Heterogeneous effects on management

	Management 2015	Principal gender	Principal tenure	Marginalization
Direct training	0.095	0.091	0.18^{***}	0.22^{***}
	(0.071)	(0.069)	(0.065)	(0.075)
Direct training × Covariate	0.070	0.080	-0.012	-0.16
	(0.071)	(0.11)	(0.0087)	(0.11)
Covariate	0.20^{***}	0.024	0.0095	0.016
	(0.053)	(0.081)	(0.0065)	(0.086)
No. of obs.	511	913	913	913
Control mean	0.12	-0.05	-0.05	-0.05

This table shows the results from estimating Equation 2.2 when the outcome variable is the 2018 DWMS index. "Management 2015" refers to the index calculated with baseline information, "Principal gender" takes a value of 1 for female principals and 0 for males, "Principal tenure" refers to the number of years as principal, and "Marginalization" takes a value of 1 for schools located in areas with high or very high marginalization. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.14: Heterogeneous effects on learning

	Student gender	Management 2015	Principal gender	Principal tenure	Marginalization
Direct training	0.033	0.046	0.0076	0.010	0.047
	(0.030)	(0.042)	(0.037)	(0.037)	(0.044)
Direct training × Covariate	0.0063	-0.052	0.056	0.0043	-0.027
	(0.025)	(0.042)	(0.059)	(0.0051)	(0.057)
Covariate	0.22^{***}	0.037	0.070^{*}	-0.00023	-0.21^{***}
	(0.018)	(0.026)	(0.041)	(0.0034)	(0.043)
No. of obs.	37,958	19,112	37,958	37,867	37,958
Control mean	-0.02	0.00	-0.02	-0.02	-0.02

This table shows the results from estimating Equation 2.2 when the outcome variable is the PCA index from math and language 2018 PLANEA scores. "Management 2015" refers to the index calculated with baseline information, "Principal gender" takes a value of 1 for female principals and 0 for males, "Principal tenure" refers to the number of years as principal, and "Marginalization" takes a value of 1 for schools located in areas with high or very high marginalization. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.15: Treatment effect on principal turnover

	Turnover
Direct training	-0.030
	(0.031)
No. of obs.	1,010
Indirect training mean	.43

This table presents the treatment effects on principal turn over (i=1 if there is a change in the school's principal between 2015 and 2018). All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.16: Heterogeneous effects of principal turnover

Panel A	Management
Direct training	0.16**
	(0.073)
Direct training × Principal change	-0.087
	(0.10)
Principal change	0.025
	(0.073)
No. of obs.	909
Control mean	-0.05
Panel B	PLANEA 2018
Direct training	0.042
	(0.039)
Direct training × Principal change	-0.027
	(0.059)
Principal change	-0.063
	(0.042)
No. of obs.	37,465
Control mean	-0.01

This table shows the results from estimating Equation 2.2 when the covariate is principal turnover (a change in the principal between 2015 and 2018). Panel A has as the outcome variable the 2018 DWMS index. Panel B has the PLANEA 2018 score as the outcome variable. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Relationship between management practices and the Stallings class- room observation and SisAT tools

Was it reasonable to expect that providing training on two tools would improve managerial practices and test scores? We address this question by looking at the correlation between the self-reported information on the use of the Stallings classroom observation and SisAT tools on both the DWMS and test scores. As mentioned above, these results should be interpreted with caution since they rely on school principals' self-reported assessments. Beyond measurement error problems, schools that answered the online surveys are statistically different from those that did not in several observable characteristics, including treatment status (see Tables A.9–A.12). Hence, this section does not attempt to establish a causal relationship between the use of management tools, Stallings and SisAT, and the DWMS or test scores. Instead, the three correlations described in this section are presented for completeness.

First, using both tools is positively correlated with the DWMS (see Table A.17). In other words, the more likely principals are to use the management tools, the higher the DWMS index. Second, "direct training" schools are more likely than those that received cascade-style training to implement both tools (see Table A.19). Thus, the "direct training" intervention was more successful than the cascade intervention at encouraging principals to use the management tools. Combining these two results—"direct training" schools are more likely to use the management tools provided to them, and these tools are correlated with the DWMS—it is unsurprising that the treatment improves DWMS scores (as shown in Panel A, Table 5). Finally, the self-reported information also shows that the correlation between using the management tools and test scores is not statistically significant (see Table A.18), which is aligned with the finding that the treatment did not improve learning outcomes (as shown in Panel B, Table 5).

Table A.17: Association between DWMS and implementation and use indexes

	(1)	(2)	(3)	(4)
	Management index			
Implementation index Stallings	0.049^{*}			
	(0.029)			
Use index Stallings		0.064^{**}		
		(0.028)		
Implementation index SisAT			0.064^{**}	
			(0.027)	
Use index SisAT				0.086^{***}
				(0.033)
No. of obs.	650	645	691	686

This table presents the conditional correlation between DWMS and implementation and use indexes. The implementation and use indexes are constructed as the simple average of the online survey variables for each element of the intervention. The management index and implementation and use indexes are standardized. All regression controls for strata fixed effects and enumerator fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.18: Association between learning outcomes and implementation and use indexes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Math				**Language**				**Average**			
Implementation index Stallings	0.024				0.017				0.024			
	(0.016)				(0.016)				(0.016)			
Use index Stallings		0.0087				0.027				0.020		
		(0.018)				(0.017)				(0.019)		
Implementation index SisAT			0.0044				0.014				0.0098	
			(0.015)				(0.015)				(0.016)	
Use index SisAT				-0.0084				-0.022				-0.016
				(0.017)				(0.017)				(0.017)
No. of obs.	27,643	27,516	28,994	28,711	27,966	27,837	29,247	28,965	26,682	26,561	28,076	27,807

This table presents the conditional correlation between learning outcomes and implementation and use indexes. The implementation and use indexes are constructed as the simple average of the online survey variables for each element of the intervention. The learning outcomes and implementation and use indexes are standardized. All regressions control for strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

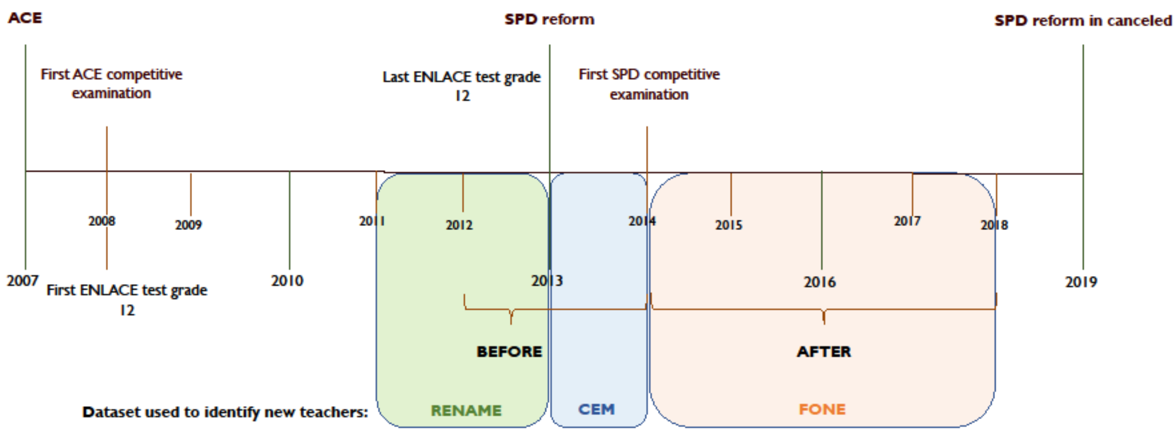
Table A.19: Treatment on Stallings and SisAT implementation and use

	Implementation index Stallings	Use index Stallings	Implementation index SisAT	Use index SisAT
Direct training	0.39***	0.38***	0.15**	-0.082
	(0.074)	(0.071)	(0.067)	(0.068)
No. of obs.	827	822	866	860
Control mean	-0.27	-0.23	-0.08	0.05

Both the implementation and use indexes are constructed as the simple average of the online survey variables for each element of the intervention. All regressions take into account the randomization design—i.e., include strata fixed effects. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B: supplemental material chapter 3

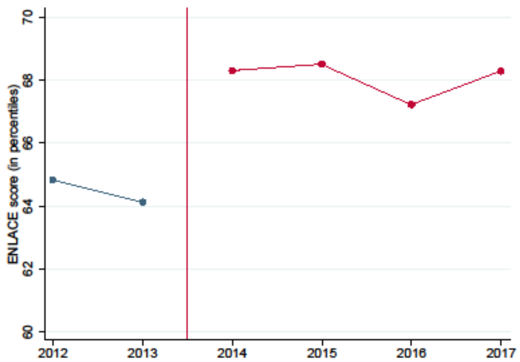
Figure B.1: Timeline and data sources



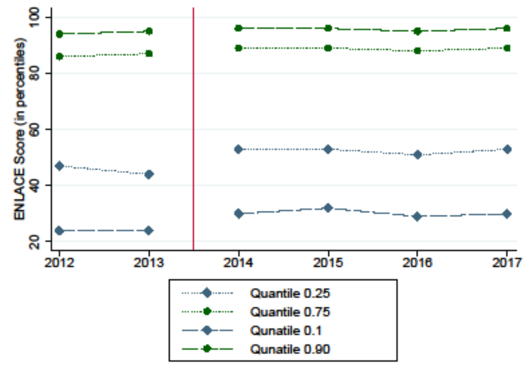
Notes: This figure illustrates in a timeline the significant changes in teacher personnel policies that the Mexican education system experienced throughout the 2007-2019 period. The figure also includes the data sources used to identify the newly hired teachers (RENAME, CEM and FONE), competitive examinations (ACE and SPD) and the ENLACE 12th grade Scores.

Figure B.2: New teachers: ENLACE Score

(a) Mean



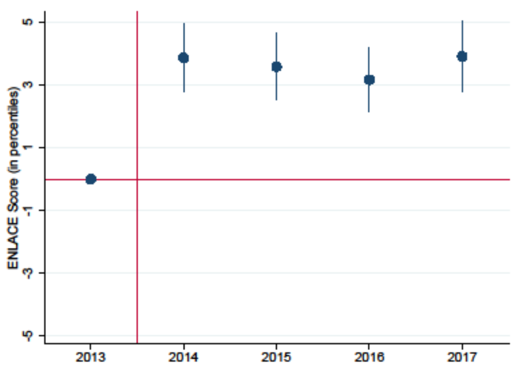
(b) Percentiles



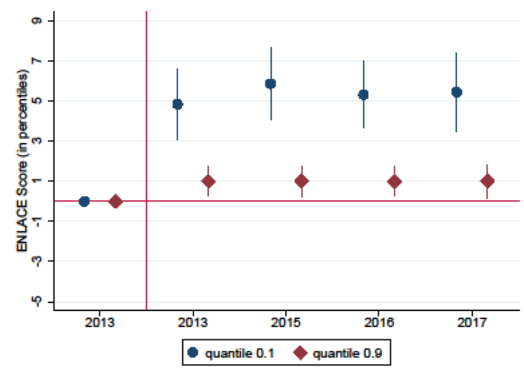
Notes: The figure (a) shows the ENLACE Score in percentiles means for newly hired teachers for each year in the sample. The figure (b) shows the ENLACE Score in percentiles at quantile 0.1, 0.25, 0.75 and 0.9 for f newly hired teachers for each year in the sample.

Figure B.3: New teachers (4 & 5 years after graduating from secondary school): Change in ENLACE scores with respect to 2013

(a) OLS estimates



(b) Quantile regression estimates

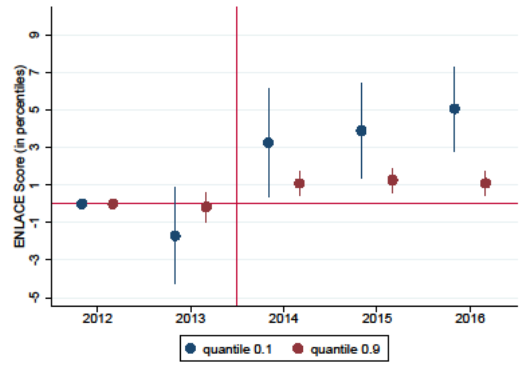


Notes: The figure (a) shows the difference in percentiles of the ENLACE Score of newly hired teachers with respect to 2013, corresponding to the βπ estimates of equation (3.1). The figure (b) shows the difference in percentiles of the ENLACE Score of newly hired teachers for quantile 0.1 in blue and quantile 0.9 in red with respect to 2013 corresponding to the βπ(0.9) and βπ(0.1) estimates of equation (2). Regressions include state fixed effects and state job market controls and robust standard errors. Confidence intervals at 95% level are shown in bars.

Figure B.4: Only 2017 FONE data: New teachers: change in ENLACE score with respect to 2012

(a) OLS estimates (mean 2012= 63.4)

(b) Quantile estimates (2012: 10th percentile = 23.5, 90th percentile =94.1 in 2012)



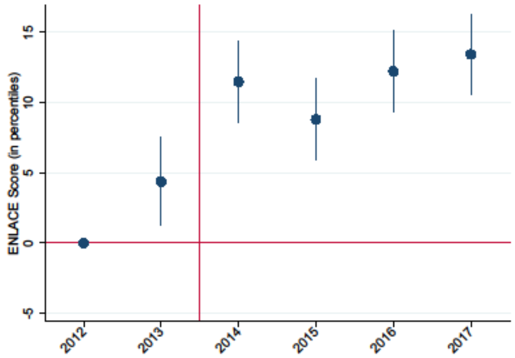
Notes: The figure (a) shows the difference in percentiles of the ENLACE Score of newly hired teachers with respect to 2012, corresponding to the βπ estimates of equation (3.1) using a FONE data set with information on the entrance date covering up to the third quarter of 2017. The figure (b) shows the difference in percentiles of the ENLACE Score of newly hired teachers for quantile 0.1 in blue and quantile 0.9 in red with respect to 2012 corresponding to the βπ(0.9) and βπ(0.1) estimates of equation (3.2) using a FONE data set with information on the entrance date covering up to the third quarter of 2017. Regressions include state fixed effects, state job market controls and robust standard errors. Confidence intervals at 95% level are shown in bars.

Figure B.5: OLS estimates: change in share of ruled-based hires (2012: mean = 67.7)

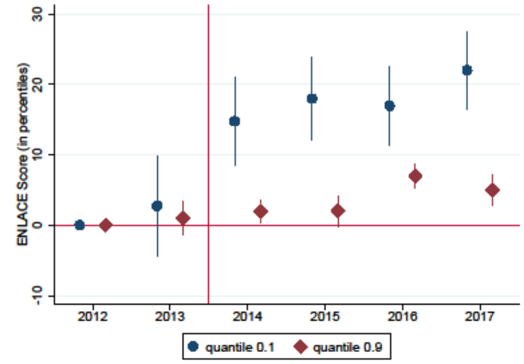
Notes: The figure shows the difference in the share of ruled-based hires within the newly hired teachers with respect to 2012, corresponding to the βπ estimates of equation (3.1). Regressions include state fixed effects, state job market controls and robust standard errors. Confidence intervals at 95% level are shown in bars.

Figure B.6: Ruled-based - Discretionary hires skills gap: change with respect to 2012

(a) OLS estimates



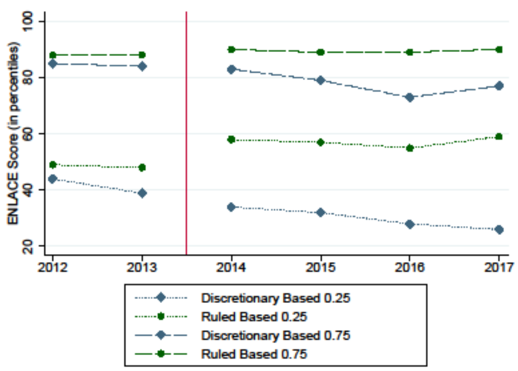
(b) Quantile regression estimates



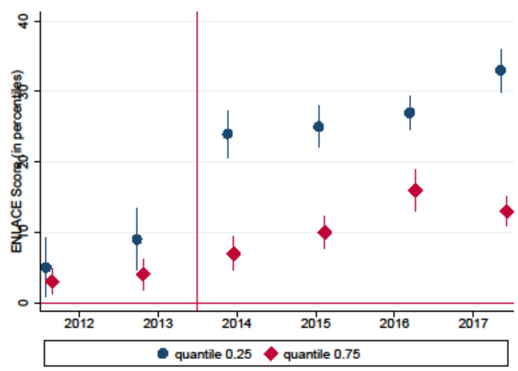
Notes: The figure (a) shows the estimates of the skills gap in percentiles of the ENLACE Score of ruled-based hires over discretionary-based hires with respect to 2012. The figure (b) the estimates of the skills gap in percentiles of the ENLACE Score of ruled-based hires over discretionary-based hires with respect to 2012 for quantiles 0.1 and 0.9. Regressions include state fixed effects and state job market controls and robust standard errors. Confidence intervals at 95% level are shown in bars.

Figure B.7: Ruled-based vs. discretionary hires: ENLACE Score

(a) 25th and 75th percentiles



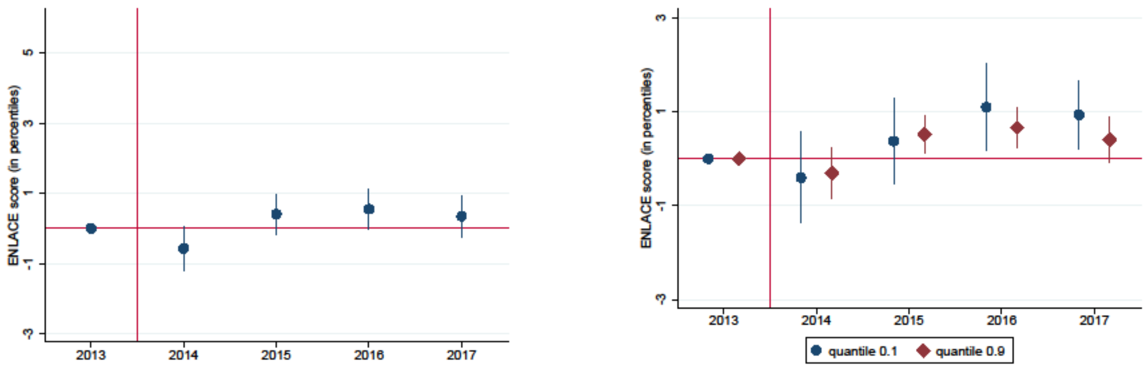
(b) Quantile regression estimates: yearly gaps



Notes: The figure (a) shows the ENLACE Score in percentiles at quantiles 0.25 and 0.75 for ruled-based hires and discretionary-based hires for each year in the sample. The figure (b) shows the estimates of the annual gaps in ENLACE Score in percentiles at quantiles 0.25 and 0.75 between ruled-based and discretionary-based newly hired teachers, regressions include state fixed effects and robust standard errors. For figure (b) confidence intervals at 95% level are shown in bars.

Figure B.8: Rule-based applicants (4 & 5 years): Change in ENLACE scores with respect to 2013

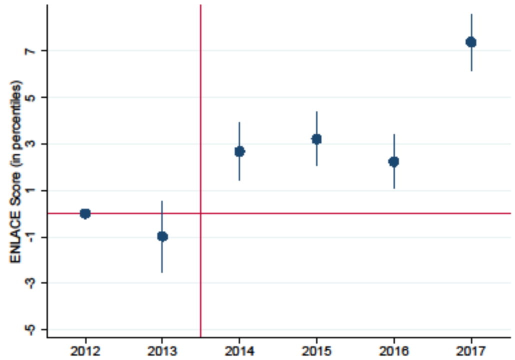
(a) OLS estimates (b) Quantile regression estimates



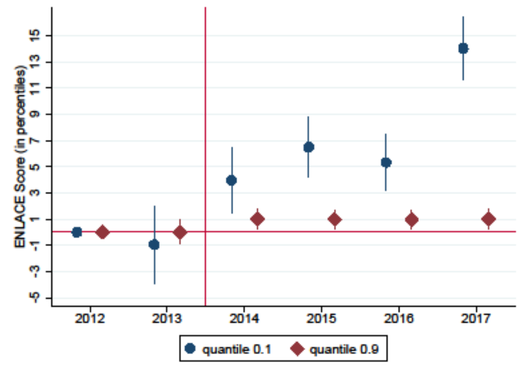
Notes: The figure (a) shows the difference in percentiles of the ENLACE Score of ruled-based applicants (4 & 5 years) with respect to 2013, corresponding to the βπ estimates of equation (3.1). The figure (b) shows the difference in percentiles of the ENLACE Score of ruled-based applicants (4 & 5 years) for quantile 0.1 in blue and quantile 0.9 in red with respect to 2013 corresponding to the βπ(0.9) and βπ(0.1) estimates of equation (3.2). Regressions include state fixed effects and robust standard errors. Confidence intervals at 95% level are shown in bars.

Figure B.9: Rule-based hires: Change in ENLACE score with respect to 2012 and 2013

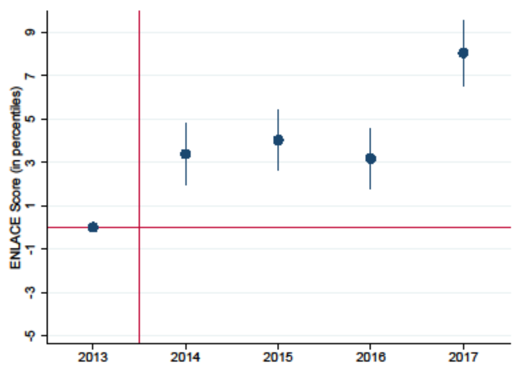
(a) OLS estimates respect to 2012



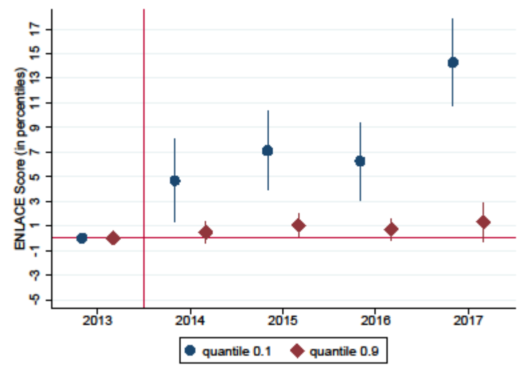
(b) Quantile regression estimates respect to 2012



(c) OLS estimates respect to 2013



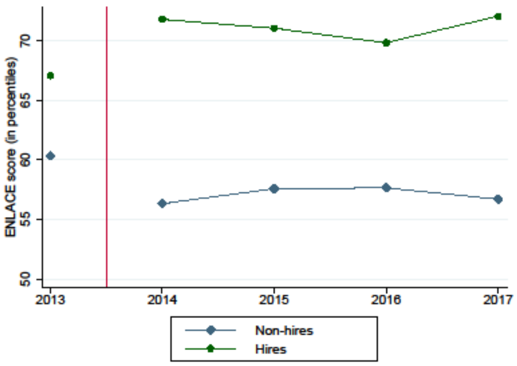
(d) Quantile regression estimates respect to 2013



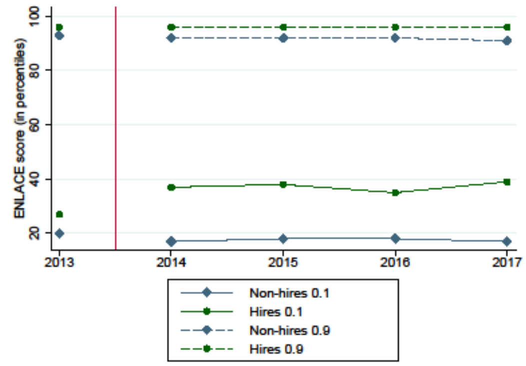
Notes: The figure (a) shows the difference in percentiles of the ENLACE Score of ruled-based newly hired teachers with respect to 2012, corresponding to the $\beta\pi$ estimates of equation (3.1). The figure (b) shows the difference in percentiles of the ENLACE Score of ruled based newly hired teachers for quantile 0.1 in blue and quantile 0.9 in red with respect to 2012 corresponding to the $\beta\pi(0.9)$ and $\beta\pi(0.1)$ estimates of equation (3.2). The figure (c) shows the difference in percentiles of the ENLACE Score of ruled-based newly hired teachers with respect to 2013, corresponding to the $\beta\pi$ estimates of equation (3.1). The figure (d) shows the difference in percentiles of the ENLACE Score of ruled based newly hired teachers for quantile 0.1 in blue and quantile 0.9 in red with respect to 2013 corresponding to the $\beta\pi(0.9)$ and $\beta\pi(0.1)$ estimates of equation (3.2) Regressions include state fixed effects, state job market controls and robust standard errors. Confidence intervals at 95% level are shown in bars.

Figure B.10: Rule based-applicants by hiring outcome: ENLACE Score

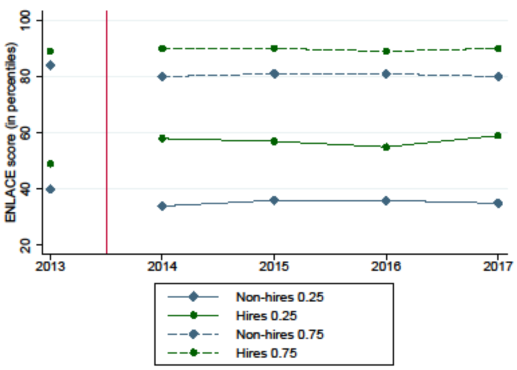
(a) Means



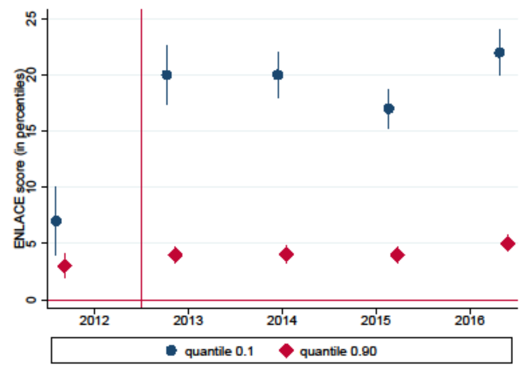
(b) 10th and 90th percentiles



(c) 25th and 75th percentiles

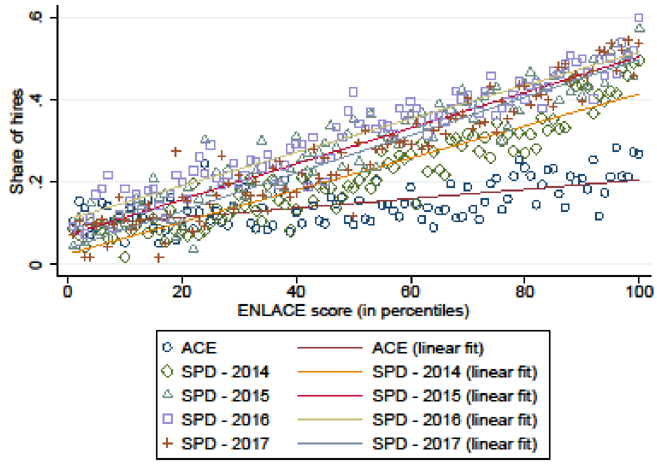


(d) Quantile regression estimates: yearly gap



Notes: The figure (a) shows the ENLACE Score in percentiles means for ruled-based Hires and Non-hires for each year in the sample. The figure (b) shows the ENLACE Score in percentiles at quantiles 0.1 and 0.9 for ruled-based Hires and Non-hires for each year in the sample. The figure (c) shows the ENLACE Score in percentiles at quantiles 0.25 and 0.75 for ruled-based Hires and Non-hires for each year in the sample. The figure (d) shows the estimates of the annual gaps in ENLACE Score in percentiles at quantiles 0.1 and 0.9 between ruled-based Hires and Non-hires, regressions include state fixed effects and robust standard errors. For figure (d) confidence intervals at 95% level are shown in bars

Figure B.11: Ruled-based applicants: Probability of being hired by ENLACE percentile and year



Notes: The figure shows probability of being hired by each percentile in ENLACE Score for ACE and SPD (yearly). A linear fit is included.

Table B.1: OLS estimates: New teachers: Change in ENLACE score with respect to 2012

	(1)	(2)	(3)
2013	-0.800	-0.781	-1.004
	(0.758)	(0.761)	(0.776)
2014	2.835***	2.969***	2.697***
	(0.619)	(0.631)	(0.629)
2015	3.131***	3.516***	3.170***
	(0.585)	(0.644)	(0.586)
2016	2.146***	2.645***	2.179***
	(0.580)	(0.696)	(0.588)
2017	3.609***	4.492***	3.835***
	(0.595)	(0.867)	(0.618)
No. of obs.	24,904	24,904	24,904
State FE	Yes	Yes	Yes
Job market level controls	No	Yes	No
Job market lagged variation controls	No	No	Yes

The table shows estimates of the β_π coefficients for equation (3.1), each column with a different set of controls. The dependent variable is the ENLACE Score in percentiles. Newly teachers incoming in the cycle 2012-2013 represent the excluded category. Job Market level control include the state unemployment rate and the average monthly income of workers with at least 13 years of education in sectors different to education services, the Job Market variation controls are the annual percentage changes of the ones just described. Robust standard errors * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Quantile regression estimates: New teachers: Change in ENLACE score with respect to 2012

	(q 0.1)	(q 0.9)
2013	-0.436	1**
	(1.878)	(0.502)
2014	4.685***	1.000**
	(1.370)	(0.437)
2015	6.964***	1.000***
	(1.264)	(0.388)
2016	5.882***	1.000**
	(1.173)	(0.389)
2017	6.717***	1.000**
	(1.319)	(0.409)
No. of obs.	24,904	24,904
State FE	Yes	Yes
Job market lagged variation controls	Yes	Yes

The table shows estimates of the β_π coefficients for equation (3.2), first column for quantile 0.1 and second column for quantile 0.9. The dependant variable is the ENLACE Score in percentiles. Newly teachers incoming in the cycle 2012-2013 represent the excluded category. The Job Market variation controls are the annual percentage changes of the state unemployment rate and the average monthly income of workers with at least 13 years of education in sectors different to education services. Robust standard errors * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

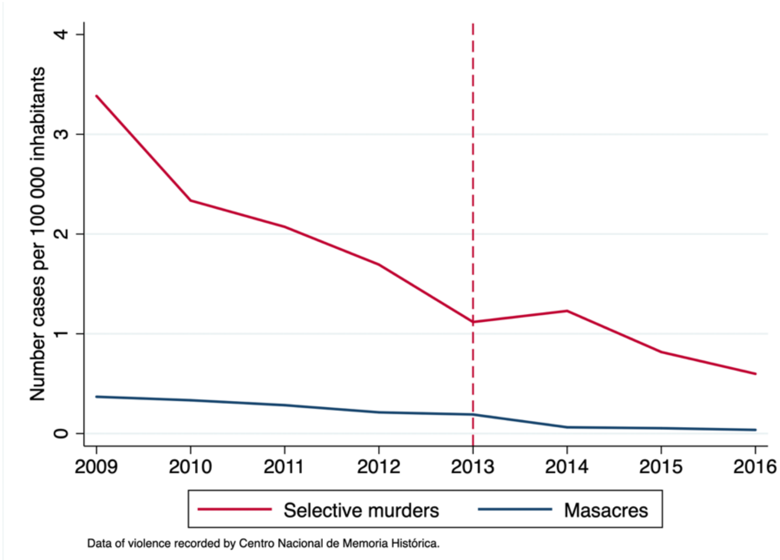
Appendix C: supplemental material chapter 4

Table C.1: Dropping Out of School

	Year						
Student	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*	*2016*
Juan	7 grade Dropping-out=0	8 grade Dropping-out=0	9 grade Dropping-out=0	10 grade Dropping-out=0	11 grade Dropping-out=0	He graduated Dropping-out=0	
Fabian	5 grade Dropping-out=0	6 grade Dropping-out=0	7 grade Dropping-out=0	8 grade Dropping-out=0	9 grade Dropping-out=0	9 grade Dropping-out=0	10 grade Dropping-out=0
Alex			4 grade Dropping-out=0	5 grade Dropping-out=0	6 grade Dropping-out=0	**He dropped out. Dropping out=1**	
Juanita				1 grade Dropping-out=0	2 grade Dropping-out=0	3 grade Dropping-out=0	**She dropped out. Dropping out=1**
Maria	4 grade Dropping-out=0	5 grade Dropping-out=0	6 grade Dropping-out=0	7 grade Dropping-out=0	8 grade Dropping-out=0	**She dropped out. Dropping out=1**	
Javier	3 grade Dropping-out=0	4 grade Dropping-out=0	5 grade Dropping-out=0	6 grade Dropping-out=0	**He dropped out. Dropping out=1**	8 grade Dropping-out=0	**He dropped out. Dropping out=1**

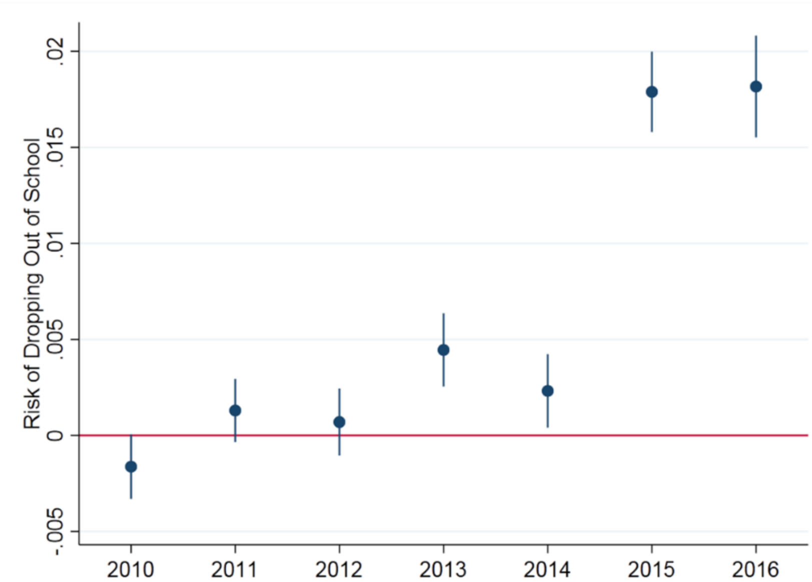
Note: This table shows how we identify school dropout. We build a dummy variable, based on information of SIMAT, which takes the value of one if the school child is not attending any educational institution and the year before he/she was enrolled in a grade lower than the graduation grade.

Figure C.2: Total Victims Nationally



Note: This figure shows the evolution of military actions associated with territorial control through fear, as selective killing and massacres of the armed groups involved in the Colombian internal armed conflict except from FARC-EP. Source: CNMH.

Figure C.3: Effect of De-escalation of the Conflict on School Attainment by Year



Notes: This figure presents the β_t coefficient from equation [4.2]. Thus, the risk of dropping out of school increased for children and youth who attended school in municipalities highly affected by FARC-EP violence, measured as those where the number of FARC-EP victims is greater than the national average before 2013.

Table C.2: Effect of De-Escalation of Conflict on School Attainment

Dependent variable: Risk of dropping out of school

	Highly exposed measure by mean				
	General [1]	without big cities [2]	Restricted to municipalities with a population in 2012 of fewer than 500,000 inhabitants [3]	Students from urban households [4]	Students from rural households [5]
CEASE x FARC	0.00933***	0.0115***	0.0124***	0.00115	0.0169***
	(0.000599)	(0.000620)	(0.000883)	(0.000733)	(0.00101)
Observations	5,510,345	4,394,027	2,120,151	3,666,590	1,843,755
Students	1063622	842563	407485	712766	350856
Households	487485	381581	181401	333707	153778
R-squared	0.216	0.215	0.216	0.218	0.217
Households FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0708	.066	.0665	.0708	.0707

Notes: This table presents the results from the main specification in equation [4.1]. *Highly exposed measure by median* is a discrete variable equal to one for municipalities where the median of the FARC-EP victim rate per 100,000 inhabitants between 2008–2012 was higher than the national median. *CEASE* is a dummy variable equal to one for the period after 2012. All regression includes sex, grade (in log), age, and age squared as controls. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. Robust standard errors area clustered at the household level. p-values for standard errors in parenthesis: * significant at 10%, ** significant at 5% and *** significant at 1%.

Table C.3: Effect of the De-Escalation of Conflict on School Attainment by Level of Education

Dependent variable: Risk of dropping out of school

Highly exposed measure by mean						
	General		Students from urban households		Students from rural households	
	Primary [1]	Secondary [2]	Primary [3]	Secondary [4]	Primary [5]	Secondary [6]
CEASE x FARC	0.00337***	0.0123***	-0.00562***	0.00392***	0.0124***	0.0224***
	(0.000947)	(0.000790)	(0.00124)	(0.000931)	(0.00144)	(0.00145)
Observations	1,951,185	3,501,130	1,182,171	2,443,549	769,014	1,057,581
Students	512559	897146	316894	624064	195665	273082
Households	314551	452860	204512	315946	110039	136914
R-squared	0.306	0.335	0.311	0.335	0.303	0.337
Households FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0545	.0777	.0567	.076	.0508	.0822

Notes: This table presents the results from the main specification in equation [4.1]. *Highly exposed measure by median* is a discrete variable equal to one for municipalities where the median of the FARC-EP victim rate per 100, 000 inhabitants between 2008–2012 was higher than the national median. *CEASE* is a dummy variable equal to one for the period after 2012. All regression includes sex, grade (in log), age, and age squared as controls. Robust standard errors area clustered at the household level. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. p-values for standard errors in parenthesis: * significant at 10%, ** significant at 5% and *** significant at 1%.

Table C.4: Effect of De-escalation of Conflict and Coca Shock on School Attainment

Dependent variable: Dropping out of school

	Highly exposed measure by mean				
	General [1]	without big cities [2]	Restrict to the municipalities with a population in 2012 less than 500,000 inhabitants [3]	Students from urban households [4]	Students from rural households [5]
CEASE x FARC	0.00299***	0.00304***	0.00492***	0.00446***	-0.000345
	(0.00102)	(0.00106)	(0.00160)	(0.00122)	(0.00185)
CEASE x FARC x COCA	0.00746***	0.0121***	0.00714***	-0.00557***	0.0195***
	(0.00128)	(0.00132)	(0.00199)	(0.00155)	(0.00226)
Observations	5,510,345	4,394,027	2,120,151	3,666,590	1,843,755
Students	1063622	842563	407485	712766	350856
Households	487485	381581	181401	333707	153778
R-squared	0.216	0.215	0.216	0.218	0.218
Households FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0708	.066	.0665	.0708	.0707

Notes: This table presents the estimation results of equation [4.3]. *Highly exposed measure by median* is a discrete variable equal to one if the median of the FARC –EP victim rate per 100, 000 inhabitants during 2008–2012 is higher than the national median. *CEASE* is a dummy variable equals one for the period after 2012. *COCA* is a dummy variable equal to one for municipalities in which their area suitable for coca cultivation is larger than the national average. All regression includes the following controls: sex, grade (in log), age, and age squared. Robust standard errors area clustered at the household level. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. P-values for standard errors in parenthesis: * significant at 10%, **significant at 5% and ***significant at 1%.

.

Table C.5: Effect of De-escalation of Conflict and Coca Shock on School Attainment by Level of Education

Highly exposed measure by mean						
	General		Students from urban households		Students from rural households	
	Primary [1]	Secondary [2]	Primary [3]	Secondary [4]	Primary [5]	Secondary [6]
CEASE x FARC	-0.00434***	0.00300**	-0.00554***	0.00402***	-0.00142	-0.000304
	(0.00166)	(0.00131)	(0.00212)	(0.00153)	(0.00267)	(0.00253)
CEASE x FARC x COCA	0.0132***	0.0106***	0.00236	-0.00104	0.0198***	0.0224***
	(0.00205)	(0.00167)	(0.00265)	(0.00196)	(0.00323)	(0.00316)
Observations	1,951,185	3,501,130	1,182,171	2,443,549	769,014	1,057,581
Students	512559	897146	316894	624064	195665	273082
Households	314551	452860	204512	315946	110039	136914
R-squared	0.306	0.335	0.311	0.335	0.304	0.337
Households FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of controls	.0545	.0777	.0567	.076	.0508	.0822

Notes: This table presents the estimation results of equation [4.3]. *Highly exposed measure by median* is a discrete variable equal to one if the median of the FARC –EP victim rate per 100, 000 inhabitants during 2008–2012 is higher than the national median. *CEASE* is a dummy variable equals one for the period after 2012. *COCA* is a dummy variable equal to one for municipalities in which their area suitable for coca cultivation is larger than the national average. All regression includes the following controls: sex, grade (in log), age, and age squared. Robust standard errors area clustered at the household level. The mean of the controls is the mean of not highly exposed to FARC violence after 2012. P-values for standard errors in parenthesis: * significant at 10%, **significant at 5% and ***significant at 1%.

Appendix D: supplemental material chapter 5

Table D.1: ICT use and student achievement

Country or territory	Observations with non-missing student controls			Observations with non-missing student & school controls		
	(1)	(2)	(3)	(4)	(5)	(6)
	δ	Se	t	δ	se	t
Albania	0,131	0,141	0,933	0,188	0,145	1,298
Brazil	0,046	0,084	0,555	0,104	0,097	1,077
Bulgaria	0,163	0,125	1,301	0,185	0,137	1,355
Chile	0,046	0,067	0,684	0,014	0,077	0,178
Chinese Taipei	-0,003	0,072	-0,043	-0,104	0,079	-1,319
Costa Rica	0,110	0,102	1,080	0,105	0,102	1,025
Croatia	0,093	0,086	1,083	0,090	0,090	0,995
Czech Republic	0,126	0,064	1,966	0,108	0,063	1,725
Dominican Republic	0,323	0,156	2,073	0,008	0,244	0,033
Estonia	0,082	0,053	1,533	0,084	0,054	1,561
Finland	-0,075	0,049	-1,538	-0,018	0,047	-0,379
France	-0,091	0,055	-1,653	-0,066	0,061	-1,082
Georgia	-0,063	0,073	-0,855	-0,053	0,077	-0,698
Germany	-0,041	0,109	-0,373	-0,100	0,125	-0,801
Greece	0,053	0,078	0,681	0,073	0,086	0,854
Hong Kong	-0,025	0,103	-0,248	0,006	0,118	0,054
Iceland	-0,013	0,046	-0,290	-0,051	0,053	-0,957
Japan	-0,126	0,072	-1,754	-0,133	0,070	-1,886
Kazakhstan	-0,022	0,147	-0,149	-0,017	0,147	-0,115
Korea	0,006	0,080	0,072	0,033	0,086	0,387
Lithuania	0,095	0,078	1,222	0,098	0,078	1,254
Luxembourg	-0,097	0,154	-0,627	-0,077	0,192	-0,402
Malta	0,218	0,308	0,707	0,238	0,317	0,751
Mexico	0,048	0,071	0,672	0,018	0,076	0,244
Panama	0,111	0,136	0,818	0,517	0,245	2,106
Poland	0,045	0,057	0,802	0,042	0,057	0,752
Russian Federation	0,005	0,099	0,050	-0,069	0,094	-0,737
Serbia	-0,182	0,107	-1,704	-0,189	0,107	-1,759
Slovak Republic	0,158	0,079	1,997	0,187	0,077	2,437
Switzerland	-0,056	0,074	-0,761	-0,044	0,074	-0,598
Thailand	0,080	0,099	0,802	0,073	0,098	0,746
Turkey	0,252	0,079	3,173	0,252	0,080	3,137
United Kingdom	0,058	0,062	0,925	0,078	0,078	0,996
United States	0,002	0,069	0,023	0,004	0,069	0,055
Uruguay	0,004	0,064	0,068	0,003	0,066	0,039
Moscow Region (rus)	0,035	0,181	0,191	0,027	0,180	0,148
Tatarstan (rus)	-0,160	0,095	-1,689	-0,155	0,096	-1,618

Table D.2: ICT use and student achievement necessary test

Country or territory	Student controls		Student and school controls	
	(1)	(2)	(3)	(4)
	Chi2	**Prob > chi2**	**Chi2**	**Prob > chi2**
Albania	16,275	0,000	13,167	0,001
Brazil	84,377	0,000	50,229	0,000
Bulgaria	65,256	0,000	60,806	0,000
Chile	76,095	0,000	18,824	0,000
Chinese Taipei	137,386	0,000	65,005	0,000
Costa Rica	129,097	0,000	81,609	0,000
Croatia	218,728	0,000	273,881	0,000
Czech Republic	183,869	0,000	95,874	0,000
Dominican Republic	120,020	0,000	72,209	0,000
Estonia	76,774	0,000	93,179	0,000
Finland	29,341	0,000	24,959	0,000
France	184,672	0,000	95,179	0,000
Georgia	11,382	0,003	17,416	0,000
Germany	20,297	0,000	39,141	0,000
Greece	334,397	0,000	269,729	0,000
Hong Kong	77,773	0,000	20,597	0,000
Iceland	1,829	0,401	2,766	0,251
Japan	4,620	0,099	2,966	0,227
Kazakhstan	1084,535	0,000	839,926	0,000
Korea	216,369	0,000	145,674	0,000
Lithuania	78,777	0,000	45,411	0,000
Luxembourg	18,035	0,000	13,053	0,001
Malta	30,408	0,000	22,002	0,000
Mexico	127,731	0,000	49,268	0,000
Panama	68,095	0,000	81,053	0,000
Poland	23,868	0,000	37,158	0,000
Russian Federation	216,014	0,000	136,805	0,000
Serbia	195,197	0,000	197,253	0,000
Slovak Republic	77,722	0,000	47,357	0,000
Switzerland	123,874	0,000	101,988	0,000
Thailand	216,926	0,000	132,669	0,000
Turkey	494,532	0,000	544,747	0,000
United Kingdom	37,840	0,000	51,415	0,000
United States	110,134	0,000	43,689	0,000
Uruguay	19,138	0,000	9,446	0,009
Moscow region (rus)	1,911	0,385	6,182	0,045
Tatarstan (rus)	69,587	0,000	110,226	0,000

Table D.3: The time of use of ICT and student achievement

Country or territory	Observations with non-missing student controls			Observations with non-missing student & school controls		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	δ	**se**	**t**	**δ**	**se**	**t**
Albania	0,0025	0,0008	3,0949	0,0024	0,0009	2,8374
Brazil	0,0002	0,0008	0,2767	-0,0002	0,0010	-0,2439
Bulgaria	-0,0005	0,0009	-0,5120	-0,0002	0,0009	-0,2359
Chile	-0,0003	0,0005	-0,7315	-0,0001	0,0006	-0,1001
Chinese Taipei	-0,0006	0,0004	-1,3506	-0,0010	0,0004	-2,3402
Costa Rica	0,0009	0,0006	1,5054	0,0009	0,0006	1,5345
Croatia	-0,0002	0,0005	-0,4009	-0,0003	0,0005	-0,5803
Czech Republic	0,0008	0,0004	1,9412	0,0007	0,0004	1,7580
Dominican Rep.	0,0015	0,0012	1,2685	-0,0005	0,0013	-0,4226
Estonia	0,0007	0,0005	1,4585	0,0007	0,0005	1,4647
Finland	0,0009	0,0004	2,4116	0,0008	0,0004	2,0578
France	-0,0011	0,0004	-2,7960	-0,0012	0,0004	-2,6553
Georgia	-0,0011	0,0008	-1,2810	-0,0013	0,0008	-1,6001
Germany	-0,0001	0,0006	-0,2629	-0,0005	0,0006	-0,8186
Greece	0,0017	0,0006	2,8526	0,0011	0,0007	1,6062
Hong Kong	0,0009	0,0009	1,0222	-0,0004	0,0007	-0,5467
Iceland	0,0011	0,0006	1,8422	0,0012	0,0006	1,9146
Japan	-0,0010	0,0007	-1,3133	-0,0011	0,0007	-1,4803
Kazakhstan	0,0019	0,0005	4,0133	0,0019	0,0005	3,9367
Korea	0,0009	0,0004	2,0361	0,0010	0,0005	1,9948
Lithuania	0,0016	0,0005	3,2710	0,0015	0,0005	3,2073
Luxembourg	0,0002	0,0006	0,3994	0,0006	0,0007	0,7826
Malta	-0,0006	0,0008	-0,7529	-0,0006	0,0008	-0,7700
Mexico	0,0008	0,0005	1,5620	0,0006	0,0006	1,0617
Panama	0,0001	0,0007	0,1736	0,0010	0,0011	0,8844
Poland	0,0010	0,0004	2,4708	0,0010	0,0004	2,5324
Russian Federation	-0,0002	0,0005	-0,3718	-0,0001	0,0005	-0,2250
Serbia	-0,0013	0,0007	-1,6833	-0,0012	0,0008	-1,5431
Slovak Republic	0,0007	0,0007	1,0786	0,0008	0,0007	1,1253
Switzerland	0,0001	0,0007	0,1372	0,0000	0,0006	-0,0527
Thailand	0,0017	0,0005	3,1263	0,0017	0,0005	3,1251
Turkey	0,0009	0,0004	2,4369	0,0009	0,0004	2,3638
United Kingdom	0,0005	0,0006	0,7251	0,0010	0,0008	1,2074
United States	0,0009	0,0004	2,3403	0,0009	0,0004	1,9916
Uruguay	0,0009	0,0006	1,4402	0,0011	0,0006	1,6552
Moscow region	0,0000	0,0010	-0,0161	0,0001	0,0010	0,0669
Tatarstan (rus)	-0,0002	0,0005	-0,3601	-0,0002	0,0005	-0,4014

Table D.4: The time of use of ICT and student achievement necessary test

Country or territory	Student controls		Student and school controls	
	(1)	(2)	(3)	(4)
	Chi2	Prob > chi2	Chi2	Prob > chi2
Albania	13,494	0,001	0,973	0,615
Brazil	49,647	0,000	57,277	0,000
Bulgaria	56,465	0,000	57,065	0,000
Chile	66,821	0,000	16,211	0,000
Chinese Taipei	38,797	0,000	18,980	0,000
Costa Rica	167,184	0,000	114,634	0,000
Croatia	152,599	0,000	211,791	0,000
Czech Republic	228,949	0,000	159,025	0,000
Dominican Republic	143,230	0,000	84,300	0,000
Estonia	53,020	0,000	66,186	0,000
Finland	58,331	0,000	44,561	0,000
France	148,787	0,000	57,696	0,000
Georgia	13,085	0,001	22,608	0,000
Germany	27,020	0,000	28,048	0,000
Greece	271,202	0,000	152,345	0,000
Hong Kong	186,631	0,000	52,908	0,000
Iceland	1,431	0,489	0,331	0,847
Japan	18,281	0,000	11,752	0,003
Kazakhstan	662,582	0,000	458,649	0,000
Korea	195,490	0,000	144,726	0,000
Lithuania	56,949	0,000	37,572	0,000
Luxembourg	129,529	0,000	18,523	0,000
Malta	25,045	0,000	22,436	0,000
Mexico	211,098	0,000	54,855	0,000
Panama	71,741	0,000	93,456	0,000
Poland	3,453	0,178	7,664	0,022
Russian Federation	132,066	0,000	114,407	0,000
Serbia	133,001	0,000	137,364	0,000
Slovak Republic	31,794	0,000	28,870	0,000
Switzerland	95,531	0,000	116,704	0,000
Thailand	89,386	0,000	62,599	0,000
Turkey	693,885	0,000	620,742	0,000
United Kingdom	74,843	0,000	145,666	0,000
United States	77,514	0,000	25,566	0,000
Uruguay	7,347	0,025	5,639	0,060
Moscow region (rus)	0,421	0,810	1,670	0,434
Tatarstan (rus)	104,370	0,000	119,322	0,000