Teachers' in-service training and student achievement:

The effect of in-service training of Peruvian teachers on student achievement

Camilo Nicanor Carrillo Purin

Teachers' in-service training and student achievement: The effect of in-service training of Peruvian teachers on student achievement

Camilo Nicanor Carrillo Purin

ISBN 978 90 8666 473 3

Publisher: Boekenplan, Maastricht

www.boekenplan.nl

Cover photo from Gisele León

Teachers' in-service training and student achievement: The effect of in-service training of Peruvian teachers on student achievement

DISSERTATION

To obtain the degree of Doctor at the Maastricht University, on the authority of the Rector Magnificus Prof. Dr. Rianne M. Letschert, in accordance with the decision of the Board of Deans, to be defended in public on Thursday 20 June 2019, at 15:45 hours

by

Camilo Nicanor Carrillo Purin

Promotors:

Prof. Dr. W.N.J. Groot

Prof. Dr. Henriëtte Maassen van den Brink

Assessment Committee:

Prof. Dr. Jo Ritzen (chair)

Prof. Dr. Frank Cörvers

Prof. Dr Inge de Wolf

Prof. Dr. Kristof De Witte

Prof. Dr. Oswaldo Molina, Lima, Peru.

Acknowledgements.

It has been 6 years, 13 visits and almost 6 months living in Maastricht during that period of time. That is the quantification and-anyone can say- the cost of a project that is coming to an end. For me was not a cost, for me these thirteen times have been a unique opportunity to visit always the Dominicanenkerk bookstore, to walk around in a city that accepted me as one of their inhabitants, a way to use this vibrant city to use it as the initial base where I started amazing trips to the rest of rest of Europe. Thirteen times that have allowed me to connect with myself and to feel that on the other side of the world there is also a place that is mine. Maastricht gave me peace, gave me happiness, and thus, was the only way I was able to fulfill a personal objective as it was this thesis.

And for that I was not alone, I received from a lot of people help, without which this long and exciting road would not have been possible. To each of those, I'm very grateful. In the first place, thanks to my promoters and supervisors Prof. Dr. W.N.J. Groot and Prof Dr. Henriëtte Maassen van den Brink. Each of our meetings on Skype, in Maastricht and even in Amsterdam were exciting, challenging and rewarding. Each recommendation and constructive criticism was key to get the best out of me. Thank you for that ability to tell me how to improve myself, trying not to stay in the easiest solution, but on the right one. You always gave me a word that made me see that I was doing well and moving in the path of a successful research.

Thanks also to the assessment committee: Prof. Dr. Jo Ritzen, Prof. Dr. Frank Cörvers, Prof. Dr. Inge de Wolf, Prof. Dr. Kristof De Witte and Prof. Dr. Oswaldo Molina. Your comments allowed me to improve this book and close the conclusions with concrete recommendations, sometimes the most difficult part for finishing a book.

Many thanks to the GPAC-Maastricht program team. During all these years I always had in Carlos, Guney, Saba and Charlotte the friends who made my stays in the seminars as comfortable and organized as possible, to allow me and my other cohorts to focus on making the best on those days. For you,

Mindel, my infinite gratitude because you always helped me making clear and predictable the whole process, but more than that, because you always did it with a big smile that made me sure that in spite of any problem I would always achieve the final objective.

And also, my thanks to whom in these years helped me with a lot of technical advice on my research. Thank you very much Micaela for so much help. We are part of a large team that in MINEDU, MEF and in the future anywhere, we have a lot to do for our country, don´t forget that.

These long 6 years filled me with enriching experiences and friends that will last forever. Corinne, Brenda, Ali, Mutinta, Paul, Clovis, Derek, Ana Cristina and all the friends of the 2013 GPAC group and other cohorts, I will always carry with me every conversation and experiences we shared together. Life will give us more opportunities to see each other, here on in the other side of the world, so see you soon. Elisa, Allison and Mary: I´ll always remember our philosophical discussions about how to survive the routine of the PhD. For Silvia and Andrea, my greatest friends during Maastricht times: here I am comrades, I finally join you in achieving this final step!

Easy? Not at all. Difficult? a lot. But who said easy things are the most exciting ones. This was probably the most difficult year of my professional and personal life, but also the most intense and the one that taught me that only with passion and effort you can reach your objectives. I was born and raised in a wonderful family where our parents taught me and my brothers everything we can do for our country and this last year only reaffirms me about that: I can do a lot to fulfill that personal and family goal.

Now is the time to continue my journey, continue with other personal and professional goals, just as difficult, just as exciting. And I will do it with the most valuable teaching that Maastricht has given me during all these years: an immense smile ...

Contents

List of Tables

List of Figures

Chapter 1: Introduction

"Those who know, do. Those who understand, teach."
—Aristotle

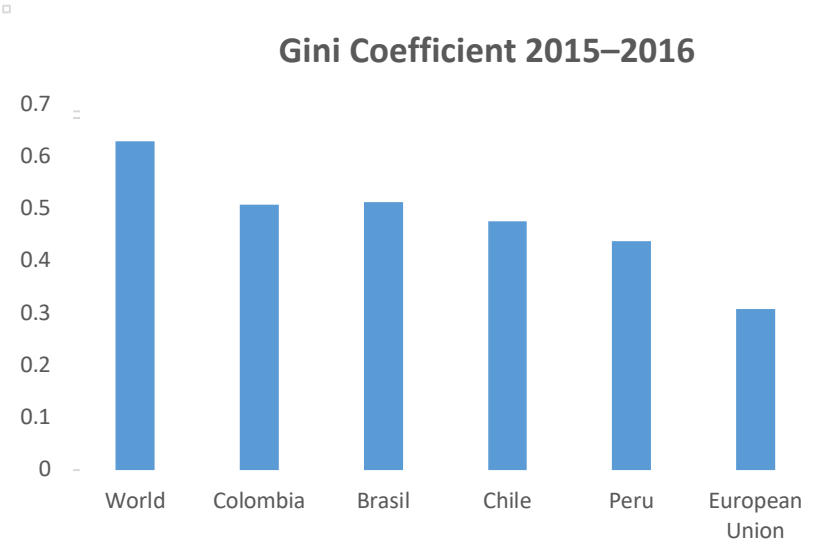
1.1 Introduction

In Peru, 2001–2017 were years of economic growth and poverty reduction. With an average annual growth rate of 5.4%, an average inflation rate of 2.9%, and a reduction in poverty from 58% to 23% over the last 15 years, the country is considered a positive example among Latin American countries (World Bank, 2017). Peru's demographic composition for its 32 million habitants in 2017 show a relatively young population with 28% below 15 years old, 65% at working age (15–64 years old), and 7% elderly habitants. However, following the global trends of developing economies, this population is facing a gradual aging process, which in the future may generate a problem for social security systems if the population does not have enough resources to cover itself, which It is directly related to their educational level.

About its geographical characterization and given its complex geography, the country can be divided in three natural regions, namely the coast (arid plains), with 55% of the population; the "Sierra" (highlands), with 32% of the population; and the "Selva" (Amazon rainforest), with 13% of the population.

Despite the relatively high average annual growth rate over the last decade, the country's economy still has room to grow, given the low per capita yearly income (US$ 12,480) compared to that of other Latin American countries in 2016, such as Chile (US$ 23,270), Colombia (US$ 13,910), and Brazil (US$ 14,810) (World Bank, 2016). Likewise, as illustrated in Figure 1 by the Gini coefficient, the income inequalities observed represent a significant challenge to Peru's future development.

Figure 1: Income inequalities in 2016: Peru vs. comparable countries and the rest of the world

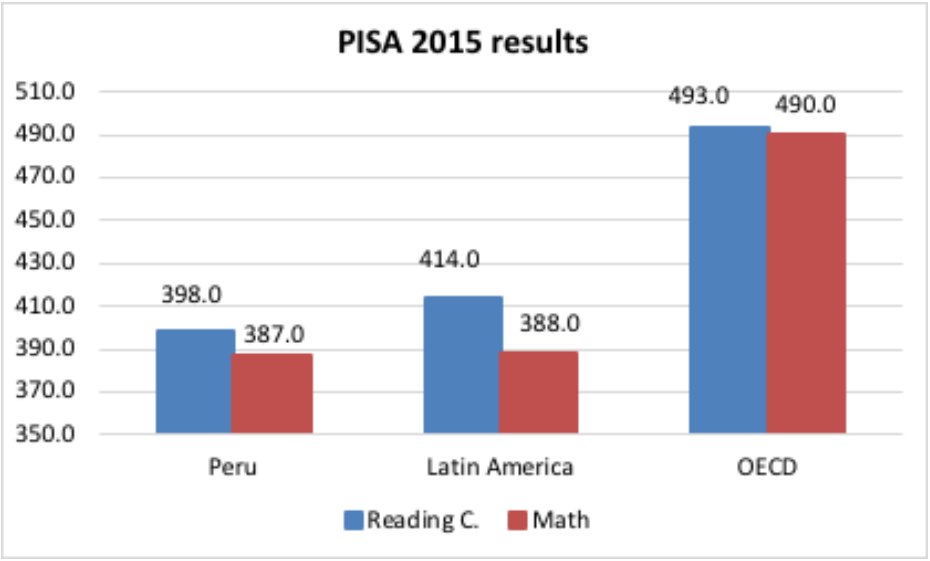


The different topographies, the levels of rurality, the youth of its population, as well as its income inequalities, make assuring high-quality educational services a challenge for the government. As Hanushek and Woessmann (2012) and Hanushek and Luque (2003) demonstrate, student achievement levels are correlated with the sustainability of long-term economic development. If the government is able to properly enhance education, Peru may be able to change from a middle-income economy to a high-income economy in the next twenty years (World Bank, 2017).

However, the incipient economic development over the last decades has not been reflected in students' academic performance. Peru ranks 67th of 75 countries/regions in Mathematics (henceforth "Math") and 68th of 75 countries/regions in Reading Comprehension in the PISA test results

framework.[1] Moreover, Peru's results are below those of its Latin American peers as well as those of the OECD countries that Peru expects to join. Figure 2 shows that for Reading Comprehension, Peru is close to the Latin American average, but far from the OECD average. For Math, the Peruvian average is below both groups. This leads to the question why the Peruvian educational model has failed to provide high-quality educational services to its current students? (Ministry of Education, 2017).

Figure 2: Comparison of PISA 2015 results for Math and Reading Comprehension



The government has not only the responsibility but also the means to improve the Peruvian educational system. Educational theory is clear in describing the supply variables, including teachers, infrastructure, and equipment, that

[1] Program for the International Student Assessment (PISA)

affect educational development and can be influenced by government policy. Researchers have determined the positive impact of teacher experience and training on student achievement (Rivkin et al., 2005; Hanushek, 2011; Mertzler and Woessman, 2010; Harris and Saas, 2011). Since 2008, the Peruvian government, within the framework of the Strategic Program for ¨Learning improvement of regular basic education students¨ (PELA)[2], has been trying to implement a professional development intervention in the form of an in-service teachers' program, providing special attention to vulnerable students (Ministry of Education-PELA Annex 2, 2015). This decision to focus on disadvantaged students in rural areas is based on an understanding of what has been proposed by Blau and Kahn (2005), Bedard and Ferrall (2003), and Checchi and van de Werfhorst (2017). These studies suggest that inequalities in educational quality affect not only students' academic results but also their future earning potential. Therefore, the focus of this thesis is on how the Peruvian professional development program for teachers, which focuses on vulnerable environments, impacts on student achievement.

This thesis evaluates a program delivered in a context of severe income inequality. Two systematic literature reviews, conducted by Kennedy (1998) and Yoon et al. (2008), have examined interventions, but only in developed economies. Therefore, this thesis fills a gap in our knowledge on the effectiveness of teacher training programs and the results can be used to aid the delivery of educational policy strategies and to reduce education inequalities in developing countries.

[2] This national program was established in 2008, and its objective is to improve the learning achievements of students from 3 to 16 years old, covering almost the entire period of public elementary education.

The Peruvian educational system, with 7.7 million basic regular students[3] in 2017, and 5.8 million of those students in public institutions, is structured into four levels. As seen in Figure 3 and described in Guadalupe et al. (2017), these four levels are:

1. Early childhood care programs (for children from 0 to 2 years old). Not compulsory, which includes an educational component. Equivalent to ISCED[4] 0;

2. Basic education. Compulsory education with three sub-levels:

    a. Initial education, for children from 3 to 5 years old. (ISCED 0);
    b. Primary education. With six grades, aimed at children from 5 to 11 years old. (ISCED 1);
    c. Secondary education. With five grades in its regular form, aimed at students from 12 to 16 years old (ISCED 2 and 3).

Parallel to the last, within this basic education there are two alternatives for those who did not have time to study at the expected age (a problem that still affects children on school age, especially in rural areas) and those who have differentiated abilities that deserve specialized attention.

This educational level according to national standards must be carried out with an approach that respects the cultural differences of the country and-if necessary-in a bilingual way for those populations whose mother tongue is different from Spanish

[3] This figure excludes students in alternative education for students who did not graduate on time (0.2 million), special education (0.02 million), and secondary technical education (0.2 million).

[4] The International Standard Classification of Education (ISCED) is a categorization used by UNESCO to compare and analyze educational levels in different countries.

3. Occupational education, intended for those who did not complete basic education and want to have a certification referring to some occupation (ISCED 2 and 3)

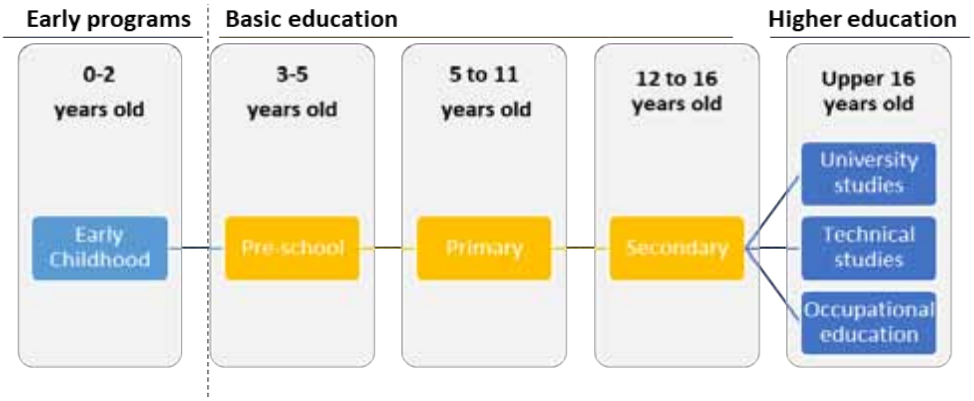
4. Higher education, which includes:

    a. Non-university education. Includes the pedagogical, artistic and technical institutes that grant a professional certification. (Equivalent to ISCED 5);

    b. University education, which offers specialized programs (ISCED 4) and professional certification of Bachelor's degree (ISCED 6), Master of Science (ISCED 7) or PhD (ISCED 8).

The teacher in-service training program developed in this thesis does not provide an academic title, so it is not included in these categories. However, within the regular programs delivered by the Ministry of Education to teachers, there is a teacher specialization program, which intervenes teachers with specialized programs in Reading Comprehension, Mathematics and English as a second language, and allows them to access to the academic degree of Master of Science.

Figure 3: Peruvian educational structure



Source: Guadalupe et al. (2017)

Multiple problems may arise as a result of low educational performance. This thesis discusses two of the important challenges that the Peruvian government will face over the next decade, namely low educational achievement in general and achievement inequalities between rural and urban students.

1. **Low student educational achievement**

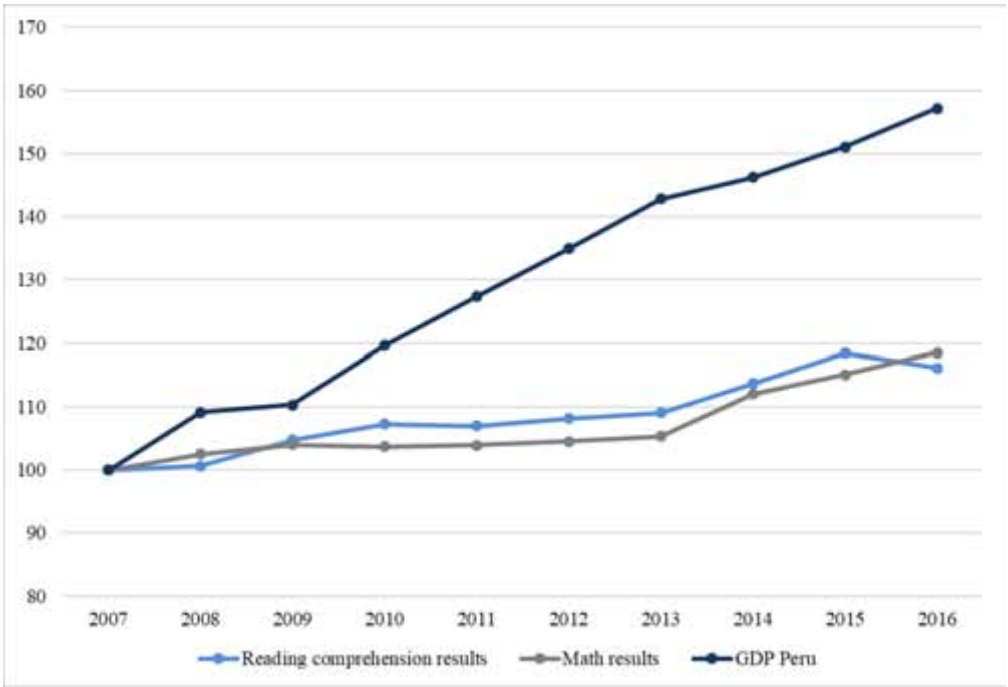
Figure 1. 4 compares the results of GDP growth to the results of the student census for each subject from 2007 to 2016. The economy during those years grew steadily, reaching an accumulated growth in GDP of 57% over a 9-year period. During the same period, average educational performance (measured in terms of knowledge compliance to respective degrees) has grown at a slower rate. Since 2015, average educational performance has stagnated at levels of 18% and 16% accumulated growth for Math and Reading Comprehension, respectively. This growth in educational returns is low,

particularly when using as a baseline the Census student assessment-ECE[5] of 2016. This baseline showed that only 34.1% of students in the second grade of primary school had achieved the grades expected in Math, and only 46.4% had performed as expected in Reading Comprehension. Although these findings are not favorable, the results are comparable with those of other countries in the region such as Chile, where a national examination of fourth-grade students in 2017 showed that 41.7% of students acquired the expected competences in Reading Comprehension, and 24.7% acquired the same skills in Math (National students results from the Education quality agency -Chile, 2018). Nonetheless, the margin for improvement is still clear.

This thesis supports the findings of Hanushek and Woessmann (2012), who state that differences in student achievement represent a significant proportion of the differences in long-term economic growth between Latin American countries and other regions of the world. For this reason, the government must prioritize policies that help the ECE graphs to have trends more similar to those in GDP growth if the government wants to ensure that the Peruvian economy continues growing in the medium and long term.

[5]The Census student assessment (ECE, by its acronym in Spanish) is a nationwide evaluation conducted annually by the Ministry of Education. The purpose of the evaluation is to obtain information on the Math and Reading performance of second and fourth graders.

Figure 4: Comparison between student results and GDP (base 100 in 2006)



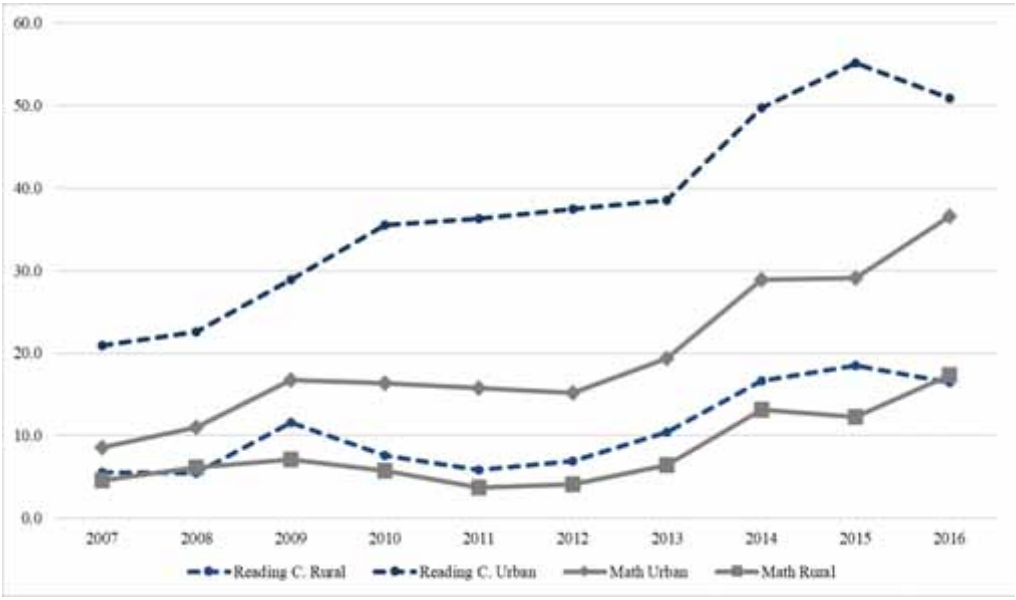
2. Inequality between urban and rural students

As observed in Figure 2, Peru has some of the lowest student achievement figures in South America[6]. In addition, when looking at local ECE 2016 average results for second-grade students reaching a satisfactory level in Reading Comprehension (46.4%) and Math (34.1%), there are notable differences when considering the gap between urban and rural students. Figure 5 compares the urban versus rural percentage of students reaching a satisfactory level for each subject between 2007 and 2016. These results show that in 2016, in Reading Comprehension, the results in urban areas (50.9%)

[6] Second lowest score of 8 Latin American countries evaluated in PISA in Reading Comprehension and third lowest score in Math.

are higher than those in rural areas (16.5%). For Math, urban areas have an attainment level of 36.6%, and rural areas have a level of 17.3%.

Figure 5: Urban vs. rural results for Reading Comprehension and Math in Peruvian second-grade students (2007–2016)



This difference is even more significant when analyzing the difference between native Spanish-speaking students and students of a different native language. According to the ECE 2016, in Math, 34.8% of native Spanish-speaking students reach a satisfactory level, whereas 23.0% of students of different native languages reach a satisfactory level. In Reading Comprehension, the percentages demonstrate similar trends (47.8% vs. 23.1% for Spanish and other native languages, respectively).

These performance gaps between Peruvian students and those from other countries, as well as the differences within Peru, may be explained by the following three factors: a low budget allocated to the educational sector; an inadequate educational infrastructure; and a deficient teacher reward structure.

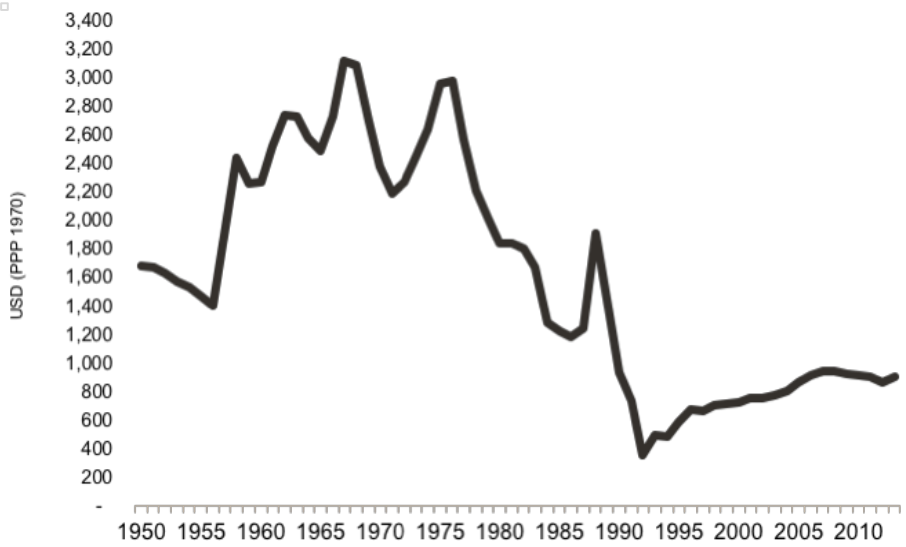
i) The educational budget: The Peruvian educational budget is only 3.7% of GDP in 2015, while the average educational budget in Latin America and Caribbean was 5.1% of the GDP (World Bank-DataBank, 2018). In terms of US$ per student, approximately US$ 1,100 is allocated to Peruvian secondary students, while US$ 2,800 is allocated to students in Chile, and US$ 2,500 in Colombia (Ministry of Education 2016).

ii) Infrastructure: The National Infrastructure Census of 2014 showed that 48% of schools must be rebuilt because they are in danger of collapsing, which means that approximately US$ 25 billion of investment is needed over the next few years[7] (Ministry of Education-Infrastructure census, 2016).

iii) Teachers' reward structure: As observed in Figure 6, teachers' salaries have decreased in real terms 47% since 1950. When comparing teachers' salaries in similar countries, Díaz, Ñopo, and Saavedra (2017) show that monthly salaries in Peru (US$ 517) are lower than those in Costa Rica (US$ 1,147), Chile (US$ 824), and Mexico (US$ 579). Also, this study shows that this difference in income is not only between countries but also within the countries' professional salaries. While in comparable economies teachers' salaries are in the 55th to 60th percentile of the national income distribution, the position of teachers' salaries is in the 35th percentile in Peru. This positioning generates a cycle where a teaching career is interesting only to people who cannot succeed in accessing better-paid professions, which prevents attracting more highly performing students to the teaching profession.

[7] Government annual investment for educational infrastructure reached US$ 1.2 billion in 2015.

Figure 6: Teachers' salaries in Peru 1950–2015 (US PPP 1970)



Source: Ministry of Education (2016)

In 2008, the Peruvian government implemented the PELA program with the objective of improving the learning achievements of students aged 3 to 16 years. The program covered the entire public school population and provided special attention to vulnerable students (who tend to be located in rural areas). The program has five areas of intervention (Ministry of Education-PELA Annex 2, 2015):

- Improvement of the educational infrastructure, which aims to improve the physical conditions where classrooms are improved through public investment in new infrastructure, as well as the rehabilitation of existing schools;

- Teachers' development, which is intended to improve the quality of teachers through specialization programs, as well as interventions that improve the development of teaching skills;

- Delivery of educational materials, whose objective is to have suitable materials for each type of student (full-grade, multigrade, bilingual) for

the development of expected learning outcomes according to a national curricular design;

- Monitoring and evaluation of educational quality, which consists of the implementation of a national system of evaluation of learning and educational quality throughout preschool, primary, and secondary education, according to the national standards;

- Evaluation of principals and teachers, which has been implemented since 2016, seeks to evaluate teachers and school directors and determine whether they meet the Ministry of Education's requirements for caring for children.

The second intervention, which is related to teacher development, includes four types of interventions intended to improve the capabilities of teachers in the classroom (Ministry of Education-PELA Annex 2, 2015):

- Curriculum management, which is aimed at preparing teachers for a more effective implementation of the curriculum;

- Teacher specialization, which updates teachers with specialized programs in Reading Comprehension, Math, English as a Second Language, and other subjects;

- Strengthening of underperforming students, which uses a set of strategies aimed to develop Math and Reading Comprehension skills in elementary students (since 2016);

- Teacher in-service training programs for regular and bilingual education.

The final intervention, the teachers' in-service program, is the professional development program assessed in this study. Established in 2008 as part of the PELA program, the program's objective is to improve, in areas of high income inequality, pre-school and primary teachers' pedagogical skills through a combination of personal classroom-based coaching, workshops,

and conferences. This program currently includes 8,726 primary schools nationwide (37% of total targeted schools in 2016) that are full-grade, multigrade, and bilingual (where students speak their native language) schools (Ministry of Education, 2016).

The program consists of three components that emphasize its pedagogical focus: (i) regular classroom visits; (ii) micro workshops; and (iii) update workshops. Classroom visits are the main form of intervention in the program. According to the Ministry of Education (2016), this component includes a first session of diagnostic observation of teacher performance by the coach and the identification of pedagogical skills that must be addressed and then continues with monthly visits to provide feedback on teachers' practices. Complementing those visits, the program includes two types of workshops, namely micro and update workshops. Micro workshops are monthly meetings between the coach and treated teachers; the coaches' objectives are to share experiences and provide comprehensive understandings of any pedagogical topics to be explained with classroom examples. Update workshops-not delivered by the coach but by a general trainer of the entire district- are meetings located outside the area of influence of the coach and include a larger group of teachers who wish to update their knowledge on general pedagogical topics.

How can this professional development program impact on student achievement? This thesis aims to address this question. To answer this question, it is necessary not only to analyze the scientific literature in a systematic review but also to examine the results from the field analysis of interventions. Within the Peruvian intervention, this evaluation must be conducted differentiating between urban and rural contexts. So, it will differentiate between urban full-grade schools, where students are divided based on their age, and multigrade schools, usually located in rural environments, where students from different ages are taught in the same classroom.

In Section 1.2, the aims of the thesis are introduced, and the research questions are developed. Section 1.3 then focuses on data description. Finally, Section 1.4 presents an outline of the thesis with a brief explanation of each chapter.

1.2 Aims and contribution of the thesis

This thesis analyses the Peruvian professional development program implemented in a context that privileges interventions in rural areas. This thesis aims to determine whether the current teachers in-service training can have a significant impact on student performance. To complement this investigation, questions about the optimal duration of this training, as well as the differentiated impacts of its components, are the main topics. The research questions are as follows:

- *Is there systematic evidence of the impact of the government's professional development program of teachers on student achievements? (Chapter 2)*
- *What is the impact of this professional development program of teachers in Peru? (Chapter 3)*
- *What is the optimal duration of the program in order to maximize its positive impact on treatment groups? (Chapter 4)*
- *Is the contribution of the professional development program's components to the positive impact on Reading Comprehension and Math homogeneous? (Chapter 5)*

The general contribution of the thesis is to further the literature on the impact of professional development programs of teachers on student achievement. Therefore, the second chapter presents a systematic literature review spanning the last 20 years (1995–2015) of research on professional

development programs around the world. In the following chapters, three different evaluations of the same professional development program in Peru are carried out. The first evaluation in Chapter 3 examines the impact of one year of intervention to determine if the program works. The second evaluation in Chapter 4 analyzes the programs where teachers have been treated for more than a year to observe marginal impacts, as the program was officially established to treat teachers for three years. Finally, the differentiated impacts of the training program components are analyzed in Chapter 5 with the aim of identifying which differentiated impacts can help in improving the efficiency of the execution of the program.

This systematic analysis of the Peruvian experience aims to offer the government empirical results that will allow policymakers to make evidence-based decisions about teacher training. These results will help to determine if it is worth implementing the program at a national level or if it is more effective to spend resources on other educational interventions that are mainly focused on rural areas.

In the following section, each research question is developed, detailing the importance and, if possible, contradictory evidence on the issue, continuing with the need to obtain an answer that provides evidence for the execution of future public policies.

Professional development programs and student achievement (Chapter 2)

Since the No Child Left behind Act[8] (NCLB, 2001) highlighted their importance, professional development programs have been seen as a way to improve teachers' skills using specialized training programs that positively affect student achievement (Yoon, Duncan, Lee, and Shapley, 2008). The first evaluations, such as Kennedy (1998), Cohen and Hill (2000), Garet et al. (2001), and Yoon et al. (2008), were published soon after the emergence of

[8] U.S. Department of Education: Title IX, Part A, (34).

these programs. These studies were the first systematic analyses of the impact of professional development programs and offered controversial evidence for their positive effect on student achievement.

Kennedy (1998) and Yoon et al. (2008) present a systematic analysis of the evidence on the impact of professional development programs on student achievement. Both papers find that a positive impact is expected when programs are intensive, content-based[9], and sustained for a longer period of time. However, the most recent studies included in Yoon et al.'s later review were published in 2006, so there is no updated information that incorporates the studies carried out during the last 10 years. Qualitative studies such as Darling-Hammond et al.'s (2009) research have suggested that teacher training must have, in addition to the three characteristics described above, clear incentives for the teacher to be effective.

However, none of the above studies have produced conclusive results. Jacob and Lefgren (2004) find conflicting evidence in Kennedy's study, in that it only found a positive impact in 12 of the 93 studies assessed. Furthermore, Chingos and Peterson (2011), using a model that controlled for school and student characteristics in Florida, show that holding a college degree with a major in education or having a Master degree is not correlated with student results in Reading and Math, which contradicts the conclusions found in Yoon et al. (2008). These inconclusive results, in addition to the lack of evidence on the impact of professional development programs in developing countries, raises questions about the expected impact of this type of training. Therefore, the topic warrants further research, particularly research that focuses on rural environments like the teacher training program conducted in Peru.

The first research question of this thesis updates the literature on the impact of professional development programs of teachers on students'

[9] Refers to programs tied to the curriculum and how teachers develop better instructional practices, not to programs focusing on teaching behaviors (Hill, Rowan, and Loewenberg, 2005).

achievements. In short, a new systematic literature review is needed. This thesis, in an effort to address that gap, uses an inclusion criterion that includes studies with experimental or quasi-experimental designs. Hence, its systematic literature review in Chapter 2 allows us to determine a causal relationship between professional development programs and student outcomes and to eliminate the bias that usually affects these programs due to the non-random assignment of the treated teachers and schools (Blundell and Costa-Dias 2009). This review also includes an analysis of the characteristics of the training programs. The objective of the analysis is to identify characteristics such as subject, duration, educational level and content of the training that might increase the probability of having positive impacts so the government can make these programs be more effective.

Professional development programs in areas with high-income inequalities (Chapter 3)

There are few causal studies on the impact of professional development programs on student achievements in developing countries that have a clear methodology and that control for the effect of teacher training from other school-based interventions. Batista Gomes-Neto and Hanushek (1994), in rural environments in Brazil, and Bando and Li (2014), in provincial public schools in Mexico, report positive and statistically significant results of in-service interventions in vulnerable environments. These findings support the development of the second research question on the impact of the Peruvian teachers' in-service program on student achievement.

The analysis to provide answers to the second research question is particularly important because the Peruvian teachers' in-service program supports pre-school and primary public-school teachers who work in vulnerable rural and bilingual institutions, following a selection criterion based on disposable income and previous student results (Rodriguez et al., 2013; Ministry of Education-PELA Annex 2, 2015). To address this second research question, an impact evaluation of the Peruvian in-service program using a statistical technique to control for selection bias of the intervention is performed in

Chapter 3. The results of this analysis sheds light on the effectiveness of the intervention and indicate whether it is a good decision to increase the scale of the program to a national level as planned by authorities. Additionally, the results contribute not only to the international literature on the impact of professional development programs of teachers but also to developing countries' efforts to increase teacher capacities through professional development programs, particularly in areas with severe income inequalities.

The impact of training duration on the professional development programs (Chapter 4)

The Peruvian teachers' in-service program trains each teacher for three years. The duration of the training is between 169–224 hours per year. This duration is specified in the Ministry of Education's program manuals, but there is no technical or substantive reason given as to why the program works with teachers for that length of time.

Looking at the impact evaluation literature, little attention has been paid to training programs and how long an optimal training program should be. Most literature refers to the duration of labor training programs, which also provides inconclusive results. Card et al. (2010), using a meta-analysis to evaluate 97 studies of employment training programs, and McGuinness et al. (2014), for job search skills and medium- to high-level skill courses, find that long-term training programs are more effective. Vooren et al. (2018) provide insights into optimal program length and state that programs that last between 7–12 months seem to be more effective in the short term, though their impact disappears in the long term. In contrast, Kluve, Schneider, Uhlendorff, and Zhao (2007), in the German labor market, find a positive impact on employment when the program lasts for less than 3 months (short-term intervention). Additionally, Hujer, Thomsen, and Zeiss (2006) find a negative effect on employment in programs that last for more than 6 months.

In terms of duration of teachers' in-service programs, Yoon et al. (2008) and Clewell, Campbell, and Perlman (2004) argue that a greater impact is expected

when programs are sustained for a longer period of time. However, only Supovitz and Turner (2000) offer an optimal training program length, finding that behavioral changes in teachers occurred after receiving 80 hours of intensive training under the Local Systemic Change (LSC) initiative[10] delivered around the USA. Beyond the findings of that study, there are no conclusive results on the optimal length of training of similar programs.

Due to the importance of determining the optimal duration of the program in order to maximize its positive impact on treatment groups, the third research question related to optimal length of the Peruvian teachers' in-service program has been developed in Chapter 4. To answer this question, a duration analysis for every year of intervention is performed. The aim is to find the optimal duration by using a comparative analysis of one, two, and three years of teacher treatment. The results of the analysis may contribute to the literature on the length of training programs, especially in the educational sector. From a local policy perspective, results will be informative for the Peruvian government in terms of determining the optimal training length, considering the scarcity of resources that Peru has for implementing educational policies.

The differential impact of components of professional development programs (Chapter 5)

Finally, upon further analysis of the Peruvian teachers' in-service program, one can observe that it consists of a clear set of interventions with three main components, namely classroom visits, micro workshops, and update workshops, each of which with different levels of exposure and characteristics. If the overall effect of the intervention is positive, with improvements in Math and Reading Comprehension, it is necessary to ask if each component contributes equally to the final results of the intervention,

[10] Teacher enhancement program financed by the National Science Foundation (NSF), which intended to improve teacher practices in Science, Math, and Technology by founding training programs based on research-based models (Supovitz and Turner, 2000).

or whether there are differentiated impacts for each one.

Then, the fourth research question, which is related to the differential effects of the in-service program components, allows this research to determine if there is room for improvement in terms of the efficiency of the program, especially in components that involve direct contact between coach and teacher during working hours, i.e., on-the-job training. Contrary to the literature that shows ambiguous results about the impacts of off-the-job interventions on educational attainment, papers like Angrist and Lavy (2001), Swinton, Scafidi, and Woodard (2012), and Bando and Li (2014) show that interventions that require on-the-job training have a positive impact on student achievement. In the labor market, Haelermans and Borghans (2011) and Mason, O'Leary, and Vecchi (2012) find that on-the-job training programs have a positive effect on employment and productivity.

Considering these findings, Chapter 5 examines the impact of the different characteristics of the training components of the Peruvian teachers' in-service program. The chapter determines if some of these components align with the definition of on-the-job training, and if so, to analyzes the differential impact of these components on educational performance.

From a policy perspective, the results of the evaluation in Chapter 5 may facilitate the Peruvian government's policies aimed at improving the efficiency of the program. The results can inform policymakers which components have no effect on learning outcomes and which do, hence indicating how to reduce, or even eliminate, certain components while increasing the frequency of those that do have an impact.

1.3 Data

Chapters 3 to 5 use cross-sectional data on primary schools in Peru. These schools participated in the Census student assessment ECE, which reports results using a "Rasch model"[11]. The results were obtained via a test designed to assess the performance level in Reading Comprehension and Math of children in the second grade of elementary school. ECE also contains variables related to school characteristics and information on other public programs (usually nutritional and cash-transfer programs) that intervene at schools. Data on the teachers who participated in the Peruvian teachers' in-service program were obtained from the information and management system for improving learning SIGMA. This system includes information on the degree of urbanity, management of schools, teachers' experiences, and progress in coaching sessions and workshops, all of which are parts of the program. This research uses also the Ministry of Inclusion's Poverty Map, which provides information about poverty, rurality, and malnutrition of territories at a district level.

Table 1 presents the relevant variables of the study, their source of information, and how they are included in the model used in Chapters 3 to 5. Student achievement ($\mathbf{Y}_{it}$) is presented as the dependent variable, school ($\mathbf{X}_{it}$) provides information on the treated school, and ($\mathbf{Z}_{i}$) includes a group of control variables that influence student achievement. The final group of variables is comprised of level of rurality; school type (multigrade or full-grade school); poverty; participation in other social programs such as JUNTOS, which is a conditional monetary incentive program; CRECER, the program aimed at improving food and nutrition practices; and the level of malnutrition.

One of the research objectives of this study is to discriminate between urban

[11] The Rasch model, usually used on educational tests, is based on the statistic methodology called Item Response Theory. This model obtains the probability of each response as a function of student abilities and the item difficulty. It´s used to measure the students results in PISA exam.

and rural environments. As such, the sample is differentiated between urban full-grade schools and rural, usually multigrade, schools (because of the scarcity of students) across the country. This sample allows the thesis to make inferences about the results of the program in both types of public schools involved.

This thesis uses different subsamples depending on each chapter's objective. The largest samples used for multigrade schools contain approximately 4,000 treated schools and 16,000 untreated schools. For full-grade schools, the largest samples are 337 treated schools and 6,000 untreated schools.

Table 1: Data variables

	Variables	Data Base Source	Description	Level
Y_{it}	Student achievement	ECE	Score	Student-school
X_{it}	Schools	SIGMA	Dummy	School
Z_1	Rural	ECE	Dummy	School
Z_2	Poverty (district level)	Inclusion Ministry Poverty Map	Percentage of poor people	District
Z_3	Malnutrition (district level)	Inclusion Ministry Poverty Map	Percentage of malnourished people	District
Z_4	CRECER (nutrition program)	ECE	Dummy	School
Z_5	JUNTOS (cash transfer program)	ECE	Dummy	School

1.4 Outline

The thesis is comprised of six chapters. Chapter 1 describes the aims and contributions of the thesis and specifies the main research questions.

Chapter 2, taking as a starting point the studies of Kennedy (1998) and Yoon et al. (2008), reviews the literature on the effects of professional development programs on student achievement over the last 20 years (1995–2015) in an effort to update the literature on the topic. The review looks for differences in the impact of the programs depending on the length of training, type of intervention (pedagogical or content-based), and degree of urbanity (urban vs. rural). Some of those aspects for the Peruvian experience, particularly training duration and intervention type, are further analyzed in Chapters 3 to 5.

Chapter 3 evaluates the impact of the Peruvian teachers' in-service program on second-grade students' achievement in Math and Reading Comprehension, expecting to find a positive impact when controlling for student characteristics. This evaluation contributes to the international review of the impact of professional development programs delivered in developing countries and contributes to the decision-making process on the program in Peru.

Chapter 4 assesses whether the positive impact of the Peruvian teachers' in-service program is sustained over time, both in Math and Reading Comprehension, when teachers receive the training consecutively for more than one year. Results will be informative in allowing the Peruvian government to decide the optimal exposure time to the program. These results can also be used to contribute to the literature on the length of training programs.

Chapter 5 assesses separately each of the Peruvian teachers' in-service program components to determine if the contribution of these components

to the overall impact of the program is homogeneous, or whether it is possible to desist from any of them and achieve the same positive impact on students. This finding may facilitate Peruvian government's policies in improving the efficiency of the program by reducing or eliminating any of the unnecessary components.

Finally, Chapter 6 presents the conclusions on the findings of the previous chapters. These findings include the policy implications, especially for governments expecting to introduce this kind of intervention in rural environments. This chapter also determines whether the intervention is cost effective in comparison with other public educational interventions delivered, not only by the Peruvian government, but also by policymakers in other countries.

Chapter 2: Professional development programs and their effect on student achievement: A systematic review of the evidence

"A teacher affects eternity; he can never tell where his influence stops."
—Henry Brooks Adams

This chapter is based on Carrillo, C., Maassen van den Brink, H., and Groot, W. (2016). Professional development programs and their effects on student achievement: A systematic review of evidence. TIER WORKING PAPER SERIES. TIER WP 16-03.

2.1 Introduction

The effectiveness of professional development programs is controversial. While some studies have found positive results (Kennedy, 1998; Yoon et al., 2008; Darling-Hammond et al., 2009), others have found negative results (Jacob and Lefgren, 2004; Chingos and Peterson, 2011). Internationally, after the No Child Left behind Act highlighted the professional development program's importance as a way of improving student achievement, and in Peru, after the PELA program implemented the teachers' in-service program based on a criterion of prioritizing intervention in rural vulnerable schools, it is clear that there is a need for further research on the subject.

Professional development is a concept used to describe systematic interventions that are part of national or local plans aimed at improving teachers' ability to understand instructional strategies based on scientific research and using technology to improve classroom development (No Child Left behind Act, 2001; Glatthorn, 1995). This method of improving teacher's skills and knowledge includes a wide range of methodologies and tools that can take the form of conferences, workshops, school-based activities, and personal coaching (Loucks-Horsley and Matsumoto, 1999).

Two important studies that have attempted to systematically analyze the relationship between professional development and student achievement are Kennedy (1998) and Yoon et al. (2008). These studies used a rigorous methodology, similar to that of the What Works Clearinghouse (WWC) Evidence Standards[12], which considers, among other concerns, the effect size of the impact on student achievement, experimental or quasi-experimental designs, and the reliability of the data.

These reviews conclude that effective interventions have the following characteristics: a focus on teachers' knowledge of the subject, sustainability

[12] US. Department of Education's standards.

over time and intensity, and an emphasis on teachers' experiences in the classroom. However, since the last systematic review of evidence by Yoon et al. (2008), which covers papers written up to 2006, there has been no study of recent findings of the literature on professional development programs, which suggests that current findings are outdated.

This chapters' objective is to review literature on professional development programs over the last 20 years (1995–2015), including studies in both developed and developing countries. These findings will contribute to the current literature and allow governments to support the implementation of teacher training programs with the ultimate goal of improving educational performance, especially in areas with high income inequalities.

The outcome measures used in this review are student achievement in Reading Comprehension and Math with a focus on primary and secondary schools. Particular attention is paid to the characteristics of professional development programs in terms of the length, training content, and degree of urbanity. Degree of urbanity has commonly been neglected in previous systematic reviews.

This chapter is structured as follows. Section 2.2 describes the search strategy and the inclusion criteria for the selection of studies included in the research. This Section includes the description of the review process and the reasons to include the chosen studies in the final list of papers.

Section 2.3 describes the characteristics of the studies found and, following a systematic review, generates information on the impact of these programs on student achievement. Special emphasis will be given to differentiating the effect of these interventions by subject, the training content, the duration and the urban-rural environment in which the program is implemented. Finally, Section 2.4 offers conclusions and discusses the main findings of the analysis.

2.2 Method

2.2.1 Search strategy

This chapter focuses on empirical studies that assess the quantitative impact of professional development programs on student achievement using a systematic approach. The search strategy of this chapter is based on existing analyses conducted on professional development programs and their impacts, such as those of Kennedy (1998), Cohen and Hill (2000), Garret et al. (2001), Fishman et al. (2003), Yoon et al. (2007), and Yoon et al. (2008), because the results of these studies are based on the use of rigorous methodology.

One key consideration in previous reviews and in this thesis as well is the inclusion of studies with high internal validity based on randomized controlled trials or quasi-experimental studies[13]. Quasi-experimental studies are important because most programs usually encounter issues of selection bias (selection of schools on the basis of students' academic or socioeconomic level (Jacob and Lefgren, 2002)) or selection of teachers to classrooms based on unobserved student or teacher characteristics (Harris and Saas, 2011).

The first stage of this systematic literature review used electronic databases to find empirical studies from the last 20 years (1995–2015) on professional development. The following databases were used: JSTOR, EBSCOhost, Education Resources Information Center (ERIC), Science Direct, and Pro-Quest. To limit the search to relevant topics, the keywords "professional development", "in-service training", "student achievement", as well as related combinations of those terms, were used. As a result of this search, 2,247 potential papers were found. The next stage involved reading the abstracts and selecting papers focused on assessing the quantitative impact

[13] As stated by Gertler et al. (2011), the most effective way to determine impact is by using randomized controlled trials. Well-implemented quasi-experimental studies are also a possibility to establish the unbiased impact of in-service training on student achievement.

of professional development programs on different outcomes. As a result, 69 potential papers were selected.

In the third stage, papers from the list of 69 potential papers, namely papers that focus specifically on the quantitative impact on student achievement, were selected. This selection reduced the sample to 12 papers. A further group of 13 papers, obtained from the selected papers bibliography, was added at this stage. As such, this stage brought this study's sample to a total of 25 papers.

In the fourth stage, the 25 chosen papers were assessed based on their ability to meet the inclusion criteria. These criteria were based on those developed by Van Klaveren and De Wolf (2014) and on the guidelines of the Campbell Collaboration.[14] The inclusion criteria thus guaranteed the quality and reliability of the information. To meet the inclusion criteria, studies had to do the following:

- Consider student achievement (Reading Comprehension, Math, Science, etc.) as the dependent variable (Inclusion criteria I).

- Have a clear and rigorous quantitative methodology that identifies the relationship between professional development and student achievement. This methodology entails the use of Randomized Control Trials (RCT) or quasi-experimental designs that show the equivalence of observables between the control and the intervention group (Inclusion criteria II).

- Include papers that have been published and previously passed a peer review (Inclusion criteria III).

[14] Non-profit organization whose aim is to give advice governments about the effects of interventions in social sectors by the use of evidence-based analysis.

- Be clear of data issues that may bias the results of the study (particularly sample size) (Inclusion criteria IV).

The following 12 papers from the previous list met the four inclusion criteria (strict sample): Angrist and Lavy (2001), Jacob and Lefgren (2004), Bressoux, Kramarz, and Prost (2008), Swinton, Scafidi, and Woodard (2012), Telese (2012), Rosangela Bando and Xia Li (2014), Harris and Saas (2011), McCutchen et al. (2002), McGill-Franzen, Allington, Yokoi, and Brooks (1999), Garet et al. (2008), Garet et al. (2011), and Glazerman et al. (2010). Additionally, a further 9 studies met all criteria except Inclusion Criteria II: Hasan Ünal, Demir, and Kilic (2011), Podhajski, Mather, Nathan, and Sammons (2009), Bacolod and Tobias (2006), Batista Gomes-Neto and Hanushek (1994), Glewwe, Kremer, Moulin, and Zitzewitz (2004), Anderson (2000), Monazza (2003), Saxe, Gearhart, and Nasir (2001), and Cohen and Hill (2000). These 9 papers are also considered on the final list of papers because they provide information on the sign and the expected impact of professional development programs on student achievement.

The four stages of the search strategy are summarized in Table 2

Table 2: Description of the systematic review process

Stages	Tool	Number of papers
First stage	Prescreening	2,247
Second stage	Reading abstract and determining quantitative impact	69
Third stage	Impact on student achievement (a)	12
	Other papers – additional analysis (b)	13
	Total (a+b)	25
Fourth stage	Inclusion criteria I,II,III,IV – Strict sample (c)	12
	Inclusion criteria I,III,IV – (d)	9
	Total (c+d)	21

2.2.3 Summary of studies included

Table 3 shows the final list of 21 papers obtained in the four-stage analysis, which includes the use of the inclusion criteria. Data are shown on the author(s), the subject (Math/Reading/other), the grades, the degree of urbanity (urban/rural), the country's level of development, the focus of training (pedagogical or content-based), the research design used, the duration of the training, the quantitative impact (with standard deviations in parenthesis), and its meeting of the inclusion criteria.

Table 3: Impact assessment studies related to professional development programs in Math and Reading Comprehension

N°	Authors	Subject	School Level	Degree of urbanity	Economic environment	Type of PD	Methodology of analysis	Duration of PD	Impact	Inclusion criteria
1	Angrist and Lavy (2001)	Math	Elementary	Urban	Developed	Pedagogy	Difference-in-difference (DiD)	240 hours	0.46* (.075)	I, II, III, IV
		Math	Elementary	Urban	Developed	Pedagogy	Propensity score matching (PSM)	240 hours	0.228* (0.082)	I, II, III, IV
		Reading	Elementary	Urban	Developed	Pedagogy	DiD	336 hours	0.616* (0.080)	I, II, III, IV
		Reading	Elementary	Urban	Developed	Pedagogy	PSM	336 hours	0.178* (0.096)	I, II, III, IV
2	Jacob and Lefgren (2004)	Math	Elementary	Urban	Developed	Pedagogy	Regression discontinuity	3 years	-0.021* (0.01)	I, II, III, IV
		Math	Elementary	Urban	Developed	Pedagogy	Regression discontinuity	3 years	0.005 (0.026)	I, II, III, IV
		Reading	Elementary	Urban	Developed	Pedagogy	Regression discontinuity	3 years	-0.020* (0.01)	I, II, III, IV
		Reading	Elementary	Urban	Developed	Pedagogy	Regression discontinuity	3 years	-0.007* (0.022)	I, II, III, IV
3	Bressoux, Kramarz, and Prost (2008)	Math	Elementary	National	Developed	Content-based	Instrumental variable	1 year	0.241* (0.091)	I, II, III, IV
		Reading	Elementary	National	Developed	Content-based	Instrumental variable	1 year	0.048 (0.077)	I, II, III, IV

N°	Authors	Subject	School Level	Degree of urbanity	Economic environment	Type of PD	Methodology of analysis	Duration of PD	Impact	Inclusion criteria
		Math	Elementary	National	Developed	Content-based	PSM	1 year	0.237* (0.101)	I, II, III, IV
		Reading	Elementary	National	Developed	Content-based	PSM	1 year	0.053 (0.085)	I, II, III, IV
4	Hasan Ünal, Demir, and Kilic (2011)	Math	Elementary	National	Developing	N/A	Analysis of variance (ANOVA)	N/A	Show means	I, III, IV
5	Swinton, Scafidi, and Woodard (2012)	Economics	Middle	National	Developed	Pedagogy	Fixed effects	16 hours	0.0571*** (effect size)	I, II, III, IV
6	Telese (2012)	Math	Middle	National	Developed	Content-based	Multiple regression	N/A	13.07* (0.66)	I, II, III, IV
		Math	Middle	National	Developed	Pedagogy	Multiple regression	N/A	0.49* (0.78)	I, II, III, IV
		Math	Middle	National	Developed	Pedagogy	T test	N/A	-3.50***	I, III, IV
		Math	Middle	National	Developed	Content-based	T test	N/A	-3.25***	I, III, IV

N°	Authors	Subject	School Level	Degree of urbanity	Economic environment	Type of PD	Methodology of analysis	Duration of PD	Impact	Inclusion criteria
		Math	Middle	National	Developed	Pedagogy	T test	N/A	-10.68***	I, III, IV
7	Podhajski, Mather, Nathan, and Sammons (2009)	Reading	Elementary	Rural	Developed	Pedagogy	T test	1 year	Show means	I, III, IV
		Reading	Elementary	Rural	Developed	Pedagogy	T test	1 year	Show means	I, III, IV
8	Rosangela Bando and Xia Li (2014)	Reading	Middle	Urban	Developing	Pedagogy	Instrumental variable	200 hours	0.73* (0.58)	I, II, III, IV
9	Harris and Saas (2011)	Math	Elementary	Urban	Developed	Content-based	Instrumental variable	50 hours	0.0030***	I, II, III, IV
		Reading	Elementary	Urban	Developed	Content-based	Instrumental variable	50 hours	-0.0002	I, II, III, IV
10	Bacolod and Tobias (2006)	Math	Middle	National	Developing	N/A	Semi parametric hierarchical regression	N/A	0.102 (0.456)	I, III, IV
11	Batista Gomes-Neto and Hanushek (1994)	Portuguese	Elementary	Rural	Developing	Content-based	Value-added modeling (VAM)	30–50 months	2.021* (t=1.98)	I, III, IV
		Math	Elementary	Rural	Developing	Content-based	VAM	30–50 months	2.111* (t=2.06)	I, III, IV

N°	Authors	Subject	School Level	Degree of urbanity	Economic environment	Type of PD	Methodology of analysis	Duration of PD	Impact	Inclusion criteria
12	Glewwe, Kremer, Moulin, and Zitzewitz (2004)	Math	Elementary	Rural	Developing	N/A	DiD	N/A	-0.023 (effect size)	I, III, IV
13	Anderson (2000)	Math	Elementary	Urban	Developing	N/A	OLS	N/A	0.157* (2.931)	I, III, IV
		Reading	Elementary	Urban	Developing	N/A	OLS	N/A	0.119* (3.117)	I, III, IV
14	Monazza (2003)	Math	Middle	National	Developing	N/A	Instrumental variable	N/A	-0.805* (-3.49)	I, III, IV
		Math	Middle	National	Developing	N/A	Instrumental variable	N/A	0.251 (0.47)	I, III, IV
		Reading	Middle	National	Developing	N/A	Instrumental variable	N/A	1.079 (1.39)	I, III, IV
		Reading	Middle	National	Developing	N/A	Instrumental variable	N/A	-0.467 (-0.5)	I, III, IV
15	McCutchen et al. (2002)	Reading	Pre-school	Urban	Developed	Pedagogy	DiD	100 hours	0.39*** (effect size)	I, II, III, IV
16	McGill-Franzen et al. (1999)	Reading	Pre-school	Urban	Developed	Pedagogy	Experimental design	30 hours	0.66 (effect size)	I, II, III, IV

N°	Authors	Subject	School Level	Degree of urbanity	Economic environment	Type of PD	Methodology of analysis	Duration of PD	Impact	Inclusion criteria
17	Saxe, Gearhart, and Nasir (2001).	Math	Elementary	Urban	Developed	Pedagogy	ANOVA	60 hours	6.17*	I, III, IV
		Math	Elementary	Urban	Developed	Content-based	ANOVA	60 hours	4.73*	I, III, IV
		Math	Elementary	Urban	Developed	Content-based	ANOVA	60 hours	4.1*	I, III, IV
18	Garet et al. (2011)	Math	Middle	Urban	Developed	Pedagogy	Experimental design	58 hours	-0.01 (effect size)	I, II, III, IV
19	Garet et al. (2008)	Reading	Elementary	Urban	Developed	Pedagogy	Experimental design	48 hours	0.08 (effect size)	I, II, III, IV
		Reading	Elementary	Urban	Developed	Pedagogy	Experimental design	60 hours	0.03 (effect size)	I, II, III, IV
20	Cohen and Hill (2000)	Math	Elementary	Urban	Developed	Pedagogy	OLS	1 year	0.1*	I, III, IV
21	Glazerman et al. (2010)	Math	Elementary	Urban	Developed	Pedagogy	Experimental design	1 year	-0.1	I, II, III, IV
		Math	Elementary	Urban	Developed	Pedagogy	Experimental design	2 years	0.2*	I, II, III, IV

N°	Authors	Subject	School Level	Degree of urbanity	Economic environment	Type of PD	Methodology of analysis	Duration of PD	Impact	Inclusion criteria
		Reading	Elementary	Urban	Developed	Pedagogy	Experimental design	1 year	0.01	I, II, III, IV
		Reading	Elementary	Urban	Developed	Pedagogy	Experimental design	2 years	0.11*	I, II, III, IV

Note: Effect size refers to the difference in means between the beneficiaries and control group divided by the standard deviation of the control group. *, **, and *** refer to being statistically significant at 10%, 5%, and 1% level of significance, respectively. N/A means “not available”.

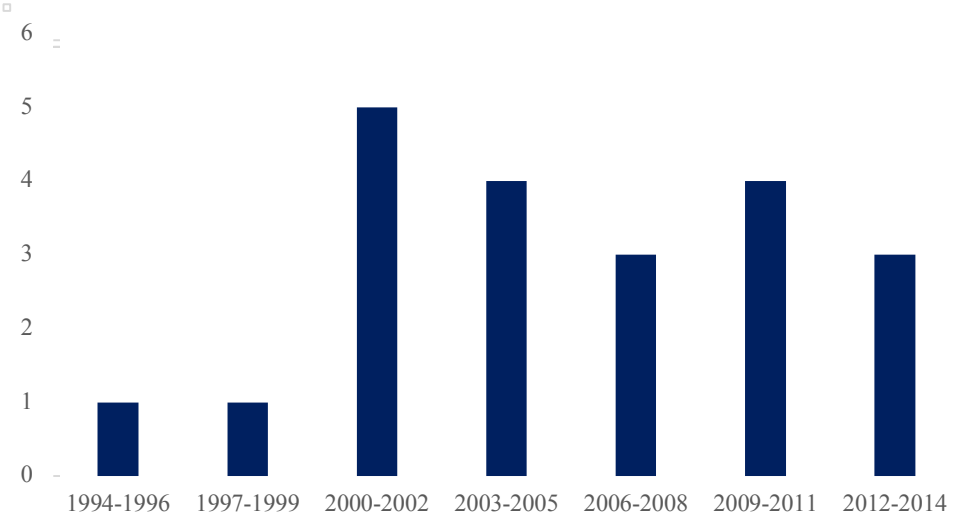
Table 4 shows that of the 21 studies, a total of 47 potentially different results are obtained (one study can include several results given their multiple analyses). 26 of these results are related to Math, and 19 are related to Reading Comprehension. In terms of school level, 32 results refer to elementary schools, 13 refer to middle schools, and 2 refer to pre-school. Most of the estimates (35) assess the impact in developed countries, and only 12 estimates study developing economies. In terms of degree of urbanity, 26 estimates focus on urban areas, 16 have a national coverage, and 5 study rural areas. Additionally, 21 estimates examine programs of over 60 hours in length, 10 study programs of 30–60 hours in length, and only 2 of them analyze programs of less than 30 hours in length. Finally, 12 estimates examine content-based programs, while 26 estimates study pedagogical interventions.

Table 4: Description of the estimates of the selected studies

	Subject			**School Level**			**Economic environment**	
	Math	**Reading**	**Other**	**Pre-school**	**Elementary schools**	**Middle schools**	**Develo-ped**	**Develo-ping**
Total	**26**	**19**	**2**	**2**	**32**	**13**	**35**	**12**
Strict Sample	12	14	1	2	20	5	26	1
	Degree of urbanity			**Duration of PD**			**Type of PD**	
	Urban	**Rural**	**National**	**Less than 30 h**	**30–60 h**	**More than 60 h**	**Content-based**	**Pedago-gy**
Total	**26**	**5**	**16**	**2**	**10**	**21**	**12**	**26**
Strict Sample	20	0	7	2	5	18	7	20

To illustrate the importance of taking studies from a new window of 20 years to ensure the systematization of the impact of professional development programs, Figure 7 presents a histogram of the years of publication of the articles in Table 3.

Figure 7:Number of papers from the four-step analysis by year of publication



As noted, half of the articles in Figure 7 were not considered by Yoon et al. (2008), whose period of analysis covers until 2005. At least 10 studies have been conducted since 2005, allowing this thesis to obtain updated data on the impact of professional development programs on student achievement.

2.3 Findings: The impact of professional development programs on student achievement

The main results obtained from the analysis of the 21 studies selected in Table 3 are summarized in Table 5. It is worth noting that, independently of the focus and characteristics of the training, 51% of the estimates provide positive and significant results, contrary to the 15% of studies that offer negative and significant results. These findings illustrate that it is likely that professional development programs have a positive impact on student achievement. About the size of the effects in the systematic analysis, for those studies that show impacts in the form of effect size, the effects range from 0.05 standard deviation (Swinton, Scafidi, and Woodard, 2012) to 0.39 standard deviation (McCutchen et al., 2002).

Table 5: Description of the study estimates by results

		Significant results		Not significant results		Total
		Positive	Negative	Positive	Negative	
Subject	Math	15	5	2	4	26
	Reading	7	2	8	2	19
	Other	2	0	0	0	2
School Level	Pre-school	1	0	1	0	2
	Elementary school	19	3	6	4	32
	Middle school	4	4	3	2	13
Degree of urbanity	Urban	15	3	4	4	26
	Rural	3	0	1	1	5
Economic environment	Developed	18	6	7	4	35
	Developing	6	1	3	2	12
Type of PD	Content-based	8	1	2	1	12
	Pedagogy	13	5	5	3	26
Duration of PD	Less than 30 hours	1	0	1	0	2
	30 to 60 hours	5	0	3	2	10
	Exceed 60 hours	13	3	3	2	21
Subject – Type of PD	Math – Content-based	7	1	0	0	8
	Math – Pedagogical	6	3	0	3	12
	Reading – Content-based	0	0	2	1	3
	Reading – Pedagogical	6	2	5	0	13

Specific results obtained from the previous table discriminating the sample in terms of subject, school level, degree of urbanity, economic environment, duration and type of professional development can be summarized as follows:

- **Subject:** There is more evidence of a positive effect of professional development programs on Math than on Reading Comprehension. 58% of the estimates that focus on student's Math performance find that the impact of teacher training is positive and statistically significant. However, only 37% of the estimates find the same positive and statistically significant results on student's Reading Comprehension performance.

- **School level:** There is more evidence of a positive effect of professional development programs in elementary schools than in middle schools. 59% of

the estimates on elementary school find a positive and significant impact. Middle school results do not display a clear trend. 30% of the estimates find positive and significant effects, but a similar percentage shows negative and significant results. In the case of pre-schools, there is not enough information to generate any evidence.

- **Degree of urbanity:** There are no differences between the effect of professional development programs in rural and urban schools. Even though there is a small difference between positive and significant impacts in rural (60%) and urban (58%) schools, considering the small sample for rural areas (5 estimates compared to the 26 in urban schools), it cannot be concluded that the impact is different.

- **Economic environment:** Similar to the degree of urbanity, there is almost the same proportion of evidence of a positive and significant effect of professional development programs in developed and developing economies. 51% and 50% of the estimates found positive and significant effects in developed and developing economies, respectively.

- **Type of professional development program:** There is more evidence of a positive effect of content-based programs than for pedagogical ones. Considering Kennedy's (1998) classification of teacher training[15], 67% of estimates that review the effect of content-based professional development programs find a positive and significant effect on student achievement. On the other hand, only 50% of the estimates for professional development programs intended to improve pedagogical skills find a positive and significant effect.

[15] Kennedy (1998) developed a classification of teacher training based on the content of the programs: (I) groups focused on teacher behaviors applying to all school subjects; (II) groups focused on teacher behaviors applying to a particular subject; (III) groups focused on pedagogical explanations of how the students learn and (IV) groups focused on teachers' learning, leaving them to develop their own teaching practices. This paper classifies groups I and II as "content-based" and groups III and IV as "pedagogical".

- **Duration of the professional development program:** Long-term professional development programs seem to have a more significant impact. 62% of estimates that assess the effect of professional development programs of over 60 hours in length find positive and significant results. This finding is 12 percentage points higher than the programs that are of less than 60 hours.

- **Professional development type and subject:** There is more evidence of a positive effect on Math using a content-based program than a pedagogical one. 88% of the estimates find positive and significant results for a content-based program, whereas 50% of the estimate find the same results when the program is focused on pedagogy. In Reading Comprehension, the opposite occurs: there is no positive estimate when the program is content-based, whereas 46% of estimates show positive results when the program is focused on pedagogical explanations.

In terms of robustness, when previous findings are compared to the analysis that only considers the strict sample (12 papers), results are similar across variables (subject, school level, degree of urbanity, type of professional development and duration of the training)[16].

2.4 Conclusion

A systematic review of literature on the effectiveness of professional development programs on student achievement has been carried out in this chapter. Independent of the characteristics of the training, 51% of the estimates find positive and significant results, allowing this thesis to conclude that teacher training programs tend to have a positive impact on student achievement.

Other findings also reveal that the probability of having positive impacts on student performance can be increased depending on certain characteristics of

[16] In the case of the variable "Economic environment", in the only study examining developing economy in the strict sample, there is a positive and significant effect. However, this result cannot be used to compare with results given by the analysis on developed economies (26), where 50% of results are positive and significant.

the programs. Firstly, positive results are more likely when the intervention focuses on improving students' Math levels than when it focuses on improving Reading Comprehension, which supports Boyd et al.'s (2009) findings on the significant impact of in-service programs on primary teachers when emphasizing Math training. Secondly, the content of the training is important, and so programs that enhance teachers' knowledge of the subject (content-based programs) show in 67% of the papers´ estimates positive results, whereas 50% of the papers´ estimates on programs intended to improve teachers' pedagogical skills have significant positive impacts. These confirmed the findings of Kennedy (1998); Hill, Rowan, and Loewenberg (2005), and Boyd et al. (2009), who argue that more significant impact on student achievement can be observed from programs that provide teachers with general guidance on the curriculum and training for a specific subject. However, when classifying the impact on the subject depending on the type of intervention (content-based or pedagogical), content-based programs have a more significant impact on Math performance, while the use of pedagogical programs has a greater impact when the aim is to improve students' achievement in Reading Comprehension. The explanation of this can be found in Carpenter et al. (1989), who find that content-based programs focused on Math provide teachers with the least information on what to do in their classrooms and instead of that provide specific information about the mathematical content to be taught and how students understand that content, which maximizes the impact on students.

In terms of the length of the training, more effective results are obtained when programs are sustained for a longer period of time (+60 hours), a finding that is supported by a significant amount of the literature on the impact of training in the educational sector (Kennedy, 1998; Yoon et al., 2007; Yoon et al., 2008; Clewell et al., 2004; Navarro and Verdisco, 2010) and the labor sector (Card et al., 2010; McGuinness et al., 2014). In spite of this body of research, none of the papers chosen in the systematic review show evidence on an optimal training duration. However, Supovitz and Turner (2000) show that for professional development programs under the framework of the Local Systemic Change (LSC) initiative, after 80 hours of intense training there are changes in the behavior of teachers.

Finally, contrary to what has been found by Podhajski, Mather, Nathan, and Sammons (2009), the degree of urbanity of professional development interventions appears not relevant to their impact on student performance. This literature review expected to find differences due to the huge inequalities in education between urban and rural areas and developed and developing countries, but the studies included in the review have found that training programs have a similar impact when implemented in different locations.

Chapter 3: The effects of in-service teacher training programs on primary-school student achievement

"Love is a better teacher than duty."
—Albert Einstein

This chapter is based on Carrillo, C., Maassen van den Brink, H., and Groot, W. (2017). The Peruvian teacher in-service training program and its effects on student achievement. TIER WORKING PAPER SERIES. TIER WP 17-14.

3.1 Introduction

The first research question developed in this study was addressed by updating the literature with the latest findings regarding the impact of professional development programs on student achievement. In addition to illustrate the greater probability of obtaining positive impacts on student performance when using these programs, the review suggests that content-based programs have a more significant impact than pedagogical ones, and that a combination of Math content-based programs and Reading Comprehension pedagogical programs may increase the probability of positive results. Furthermore, although there are no data on the optimal training length, the findings show that better results are obtained when programs are sustained for a longer period of time (+60 hours). In terms of the differing impact on urban and rural environments, this overview does not provide conclusive results due to the lack of data in Chapter 2 on rural programs.

This lack of conclusive information on the urban/rural impact is a more pressing issue when public policy advice is needed for governments, such as that in Peru, who are investing in large-scale rural intervention, which occupies a key component of their educational budgets. After a decade prioritizing enrollment rates with considerable success, with rates from 54% to 88.4% of coverage for pre-school, 93% to 93.4% for primary schools, and 69% to 83.0% for secondary school from 2001 to for 2016[17], the educational sector's priorities in Peru have changed. In 2016, 58% (US$ 4.8 billion) of the Peruvian educational budget was given to the PELA Strategic program. This budget included the growth of the PELA's in-service training interventions, which during 2016 covered 8,726 schools nationwide. These interventions have been implemented without any quantitative impact evaluation.

[17] These findings consider attendance at the official age and do not consider those who study later.

In Peru, there is a consensus on the importance of teacher training programs as a way to improve student performance in multigrade (usually rural) schools, particularly when the national student achievement results vary by almost 37 percentage points compared to full-grade schools in Reading Comprehension and 17 percentage points in Math (ECE, 2015). However, because there was no quantitative analysis for the 2009–2012 period, when the focus of the in-service training was different, this consensus is not based on an analysis of the evidence from similar programs in Peru or comparable countries. As Rodriguez, Sanz, and Soltau (2013) point out, the in-service program has been in operation since 2008 without a logical framework that would allow an evaluation of its design and implementation. Therefore, it is impossible to judge its effectiveness.

This chapter's objective is to assess the impact of the Peruvian teachers' in-service professional development program on second grade students' achievement in Math and Reading Comprehension across multigrade and full-grade public schools. These subsamples allow this chapter to differentiate the program's impact across a range of contexts. The results may determine the effectiveness of the intervention, and whether it is a good decision to increase the scale of the program to the national level (nowadays it only covers part of the potential population). Additionally, the results will contribute to the empirical evidence already collected and described in Chapter 2, particularly studies on interventions implemented in areas of high income inequality.

Data from second graders, including results, student characteristics, and school characteristics are available because the Peruvian Ministry of Education has been collecting them annually since 2007 with the ECE, the Information and management system for learning improvement (SIGMA), and with the Ministry of Inclusion's Poverty Map (see also Chapter 1).

By using government data from 2014 and propensity score matching to account for differences in program enrollment, this chapter expects to find that the teacher training program has a positive impact on student achievement when controlling for the characteristics of the student environment, such as rurality, poverty, nutrition level, etc.

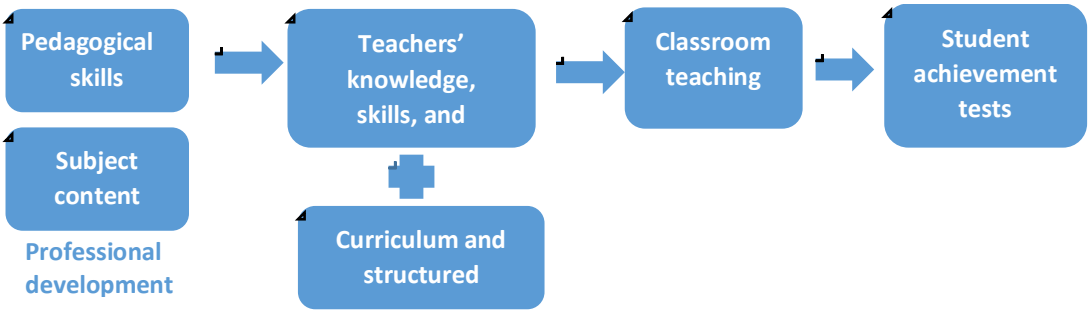
This chapter is structured as follows. Section 3.2 describes the theoretical framework of in-service professional development programs, which includes a description of similar programs delivered in North and South America. Section 3.3 describes the Peruvian teachers' in-service program. Special attention is given to the background of the program and the targeting criteria in the selection of beneficiaries, since this will be relevant for the definition of the evaluation methodology. Section 3.4 outlines the data and descriptive statistics. Section 3.5 discusses the design and empirical results, while testing for robustness. Section 3.6 concludes the chapter.

3.2 Literature

To understand the impact of professional development programs on student achievement, it is first necessary to understand the causal relationship between teachers' knowledge acquired through professional development and student results. This research uses a model based on Yoon et al. (2008) [18] (see Figure 8), which is consistent with a number of existing models (Kennedy, 1998; Cohen and Hill, 2000; Garet et al., 2001; Fishman et al., 2003; Guskey and Sparks, 2004; Bando and Li, 2014). This model assumes that the effects of professional development programs on student performance are mediated by how teachers incorporate knowledge (content-based or pedagogical skills) and reproduce them for students (Yoon et al., 2008).

[18] This process is based on Kennedy (1998) proposal on how the in-service programs affect the student achievement.

Figure 8: Professional development training logical model



Based on Yoon et al. (2008)

Figure 3.1 suggests that professional development may affect student performance in three ways. Firstly, professional development programs in the form of in-service training programs, focusing either on improving pedagogical skills or subject content, improve teachers' knowledge, skills, and motivation. Secondly, mediation through a clear curriculum and structured process, better knowledge, skills and motivation improve teaching. Third, improved teaching generates higher results on student achievement tests.

As noted in Figure 8, the only way to guarantee the final result is to assume that the training will have a direct effect on teachers' knowledge and behavior at an individual level before having an impact on classroom teaching practices. Caulfield-Sloan & Ruzicka (2005), Tienken (2003), and Borko (2004) argue that it is possible to observe the relationship between teacher training and future behavior. Additionally, Araujo et al. (2016) for Ecuadorian schools, discovered using the Classroom Assessment Scoring System (CLASS), that one standard deviation in teacher classroom effectiveness has an impact of 0.11 standard deviation on student outcomes in Language and Mathematics. Finally, Metzler and Woessmann (2012), using a Peruvian sixth-grade dataset and a within-teacher within-student variation, found that one standard deviation in subject-specific teacher achievement has a positive impact on student achievement of approximately 10% of a standard deviation, and so there is a positive causal

effect between teacher subject knowledge and student achievement. Therefore, we expect to observe that teaching skills and better teaching practices are related to students' results.

To understand the Peruvian government's decision to implement the program, it is first necessary to discuss experiences with teacher's in-service programs previously executed in the USA, Latin America, and Peru that have served as example to the Peruvian government. Firstly, the American Coaching and Mentoring for Preschool Quality (CAMP Quality) implemented, from 2010 to 2011, an eight-month professional development program designed to improve teachers' interactions with children through video-based teacher self-reflection, workshops, peer coaching, and expert coaching. Zan and Donegan-Ritter (2014) suggest that this method has a positive impact on teacher-child interactions, as measured by the Classroom Assessment Scoring System (CLASS).

Secondly, the Coaching Teachers for Emergent Literacy Program, developed in Illinois, USA, intended to deliver skill-focused coaching to certified teachers from public prekindergarten schools for one year. McCollum, Hemmeter, and Hsieh (2013) found that this program had a positive effect on teachers' use of vocabulary and language comprehension skills.

Thirdly, the Ecuadorian "Si Profe" program, introduced in 2008 in primary and secondary schools, included an annual educational support of 48 hours (eight 6-hour sessions) on topics related to improving learning methodologies, pedagogical leadership, and teacher feedback (Rodriguez, Sanz, and Soltau, 2013). Even though it was planned to expand this educational policy to the whole country, there was no independent qualitative or quantitative evaluation of its effects on student performance.

Lastly, the in-service teachers program implemented in the Brazilian state of Ceará during 2015 that combined the use of support materials, personal coaching and workshops that gave teachers benchmark information about better teaching practices to improve their classroom effectiveness. Bruns et al. (2018) using a randomized evaluation obtains that this program was able to raise 1st-year secondary student results by 0.081 standard deviation in Math and

0.047 standard deviation in Portuguese in the Ceará state student assessment.

About this type of programs in Peru, the first one was the ¨Comprehensive Teacher Training Program (Aprendes). Implemented from 2003 to 2009, Aprendes was an in-service teacher' program implemented in two poor Peruvian regions (San Martin and Ucayali) and included classroom-based coaching, periodic workshops, teachers' meetings to share experiences, and visits to similar schools. Focused on rural multigrade primary schools, the program benefitted over 16,577 students and 811 teachers. An assessment conducted between 2004 and 2008 showed positive effects on students in the first, third, and sixth grade in Reading Comprehension and Math (Bernhaum et al., 2010).

Finally, another educational policy implemented in Peru was the ¨Project for the Improvement of Basic Education¨ (Promeb), which was carried out in Piura during 2003. Funded by the Canadian International Development Agency (ACDI), the program included personalized support to teachers and principals of selected schools, providing them with continuous training on pedagogical techniques to improve their teaching practices. Alcazar and Guerrero (2011) determined the program's positive impact on third and sixth-grade students in Reading Comprehension on the ECE exam. Although this evaluation was not carried out controlling for other educational variables, the improved academic achievement during 2003–2005 may be a sign of the positive impact of the program.

3.3 Teachers' Professional development program

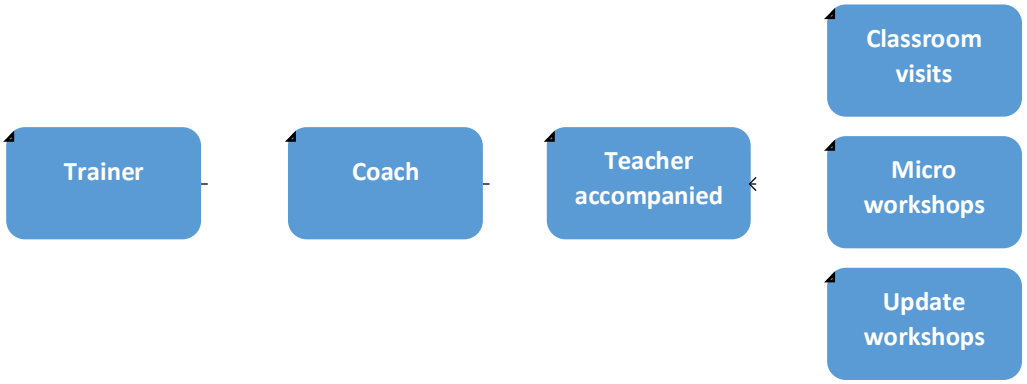
The Peruvian "in-service training", as part of the PELA, is a national program whose objective is to improve the pedagogical practice of the teacher through a systematic and permanent process of collaborative accompaniment between the coach and the intervened teacher. (Ministry of Education- In-service Program Protocol, 2014). It includes a training sequence (Figure 9) that starts

with trainers, a group of specialized personnel[19] in charge of a territorial circumscription and responsible for training coaches and providing them with technical and pedagogical assistance. Trainers should be educational professionals with experience in teacher training (Rodriguez, Sanz, and Soltau, 2013). During the first years following the implementation of the program, the selection of trainers (usually from local universities) was the responsibility of regional governments. However, after two years of negative experiences with the non-homogeneous quality of trainers, the Ministry of Education decided in 2012 to manage all contracts itself in order to have a similar baseline. As a result, regional governments were responsible only for the next stages of the programs: the selection, financing, and monitoring of coaches.

To be selected as a coach, one must meet a profile that includes certified teaching experiences in full-grade and multigrade classrooms. Additionally, in bilingual institutions, knowledge of the students' native language is also required. Teachers will receive at least 140 hours of training per year (with distance learning of approximately 220 additional hours) and this is done in three workshops with a duration of 6 days each for full-grade schools coaches and 5 days each for multigrade ones. Training includes certification of adult learning methodologies or educational support (Rodriguez, Sanz, and Soltau, 2013).

[19] Around 177 nationwide for 2014.

Figure 9: Training sequence to achieve learning outcomes



Based on update reports from the PELA (2013– 2016)

According to the protocol for 2013–2015 summarized in Table 6, the process continues with pre-school and primary teachers being trained by coaches, whose role is to provide pedagogical support (1 coach/15 teachers in full-grade and 1 coach/8 teachers in multigrade schools (Ministry of Education (2012)). The program is based on a three-year systematic approach containing three types of activities, which together entail 169–224 training hours per year. These training hours must be completed by each teacher for the teacher to be considered "fully trained" (Ministry of Education-PELA Annex 2, 2015):

1. **Classroom visits**, monthly visits where the coach observes the teacher during lecture hours, intending to give feedback that improve the teachers' skills. For the training, the coach uses an observational guide and a field notebook to record teachers' behavior in class.

2. **Micro workshops**, consisting of monthly scheduled meetings between the coach and a group of teachers to discuss more effective pedagogical practices and to stimulate reflection on classroom experiences, according to their own social and cultural context.

3. **Update workshops**, given by trainers two to three times a year consisting of specialized training for teachers to update their knowledge linked to various pedagogical topics of interest.

Table 6: Frequency and characteristics of in-service program activities (2013–2015)

Activities	Frequency (yearly)	Contact (coach vs. teacher)	Training[20]
Classroom visits	Multigrade: 9 days (72 hours) Full-grade[21]: 7 days. (35 hours)	Direct: One to one	Feedback on teacher performance: critical perspective (specific skills training)
Micro workshops	Multigrade: 8 meetings (32 hours) Full-grade: 7 meetings (14 hours)	Direct[22]: Multigrade: 1 to 8 Full-grade: 1 to 15	Share experiences and discuss pedagogical practices (specific skills training)
Update workshops	Multigrade: 3 meetings (120 hours) Full-grade: 3 meetings (120 hours)	Indirect: With the trainer and not with the coach.	Update knowledge linked to various pedagogical subjects (general training)

Source: Ministry of Education (2016)

In terms of the program's completion rate, according to SIGMA for 2015 full-grade school teachers received on average 100% of the planned classroom visits, 98% of micro workshops, and 62% of update workshops. For multigrade schools, teachers received on average 98% of classroom visits, 100% of the

[20] According to McGuinness (2014), the training components can be categorized into three groups depending to the content and focus of the program: i) previous work training; ii) general training; and iii) specific skill training. The first two groups do not have a specific linkage to the particular occupation, whereas the final group does.

[21] The frequency of the full-grade elementary schools is considered. High schools' frequency increases.

[22] It is considered "Direct" because the coach maintains contact with the teachers under his responsibility.

micro workshops, and 65% of update workshops. By 2015, a total of 23,668 teachers and 11,176 schools were treated by the program, which provides sufficient data to determine the impact of the intervention.

Finally, and necessary in the methodology of the impact evaluation, is the description of the decision-making process in allocating training. As stated by Rodriguez et al. (2013) and the Ministry of Education-PELA Annex 2 (2015), prioritization has changed in recent years. During 2008–2012, priority was given to urban full-grade schools where ECE results were lower. Since 2013, the focus has shifted towards rural and bilingual institutions. There, recent selection, and that relevant for the present analysis, is based on the following criteria:

a) Equity: schools located in districts with extreme poverty (income quintiles 1 and 2[23]);

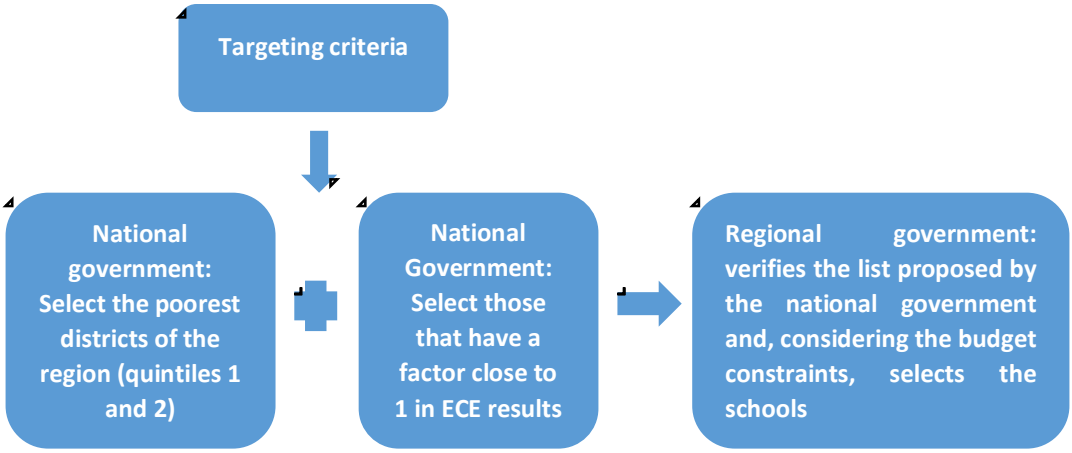
b) Student achievement: districts with low levels of learning achievement on the Student Evaluation Census (ECE); and

c) Multigrade schools: multigrade, monolingual, and bilingual schools, which are usually located in rural areas.

To determine the final list of schools, first the Ministry of Education selects potential schools to be included in the training program based on the above-mentioned criteria. Next, regional governments, together with their Local Education Management Offices (UGEL), verify this information and establish a potential list. Finally, Regional governments receive funding, and based on budget restrictions and a discretionary decision, select the schools that will receive intervention (See Figure 10).

[23] Income quintiles defined by INEI (Statistics National Institute).

Figure 10: Process of selecting the targeted schools



Based on update reports of the PELA (2013–2016).

In analyzing this specific in-service training program, there are qualitative studies on the program's organization and delivery of the training like Rodriguez, Sanz, and Soltau (2013) and a quantitative analysis of the impact on students' results during 2009– 2012, when the program focused on urban poor schools with low ECE results. This impact evaluation was conducted by the Ministry of Economics and Finance in 2013 and found a positive and significant impact of 0.2 standard deviation in the average Rasch score in Math and 0.1 standard deviation in Reading Comprehension (Ministry of Economics, 2013). However, since the Peruvian government's priority shift in 2013 from urban (full-grade) schools to rural and bilingual institutions, the above results should only be considered as referential in comparison with the results of the new impact evaluation that will be developed considering the new structure of the program.

3.4 Data and descriptive statistics

To conduct the impact evaluation, this study uses data from the following sources:

1. The census student assessment (ECE), which reports second graders' results using a "Rasch" model. Scores are classified into three levels of achievement (Ministry of Education-PELA Annex 2, 2015):

- *Beginner level*, which means that the student did not reach the expected level of learning and struggles to answer the easiest questions of the evaluation;
- *In progress*, where the student does not achieve results expected for the grade but is in the process of achieving these grades; and
- *Satisfactory*, which means that the student achieved the results expected for the grade.

The ECE also includes variables related to school characteristics, such as whether the school is full-grade or multigrade, and whether other public programs have intervened at the school (such as CRECER, a nutrition program, and JUNTOS, a cash transfer program).

2. Information and management systems for learning improvement (SIGMA), which provides data on the school treated by the in-service program. This data includes information about the progress of the coaching sessions and workshops for each teacher involved in the program.

3. The Ministry of Inclusion's Poverty Map, with information about poverty, rurality, and malnutrition at a district level.

Using the sources of information described above and summarized in the variables detailed in Table 1 in Chapter 1, it is possible to define the treated sample as public schools with teachers who receive in-service training, and the control group as public schools who did not receive the same training. For 2014,

it was possible to obtain information from 4,389 treated and 25,002 control schools. Table 7 presents the descriptive statistics of both the treated and control groups on output and control variables such as rurality, poverty, extreme poverty, malnutrition, and multigrade schools.

As observed in Table 7, for Math and Reading Comprehension, the means of student results are statistically similar (P-Value > 0.05 who indicates weak evidence against the null hypothesis of equivalence), meaning that both the treated and control groups obtain similar results on ECE tests. However, for the control variables shown in Table 7, 90% of treated schools are located in rural areas, which is higher than the percentage of rural schools in the control group (71%). Poverty and extreme poverty illustrate similar trends. In a country with high levels of poverty (20.7% in 2016) and inequalities between urban/rural areas[24], Table 7 confirms the differences between the treated and control group schools. In both groups, the differences are positive and significant (60% poverty and 30% extreme poverty in the treated sample versus 54% poverty and 24% extreme poverty in the control group). In terms of nutrition levels, there is a higher level of malnutrition (40%) in treated schools than in the control group (35%). Finally, in terms of school type (multigrade/full-grade), the intervention prioritizes multigrade schools. In all the control variables the P-Value ≤ 0.05 reject the null hypothesis of equality between the control variables and increase the need for an appropriate technique that matches the samples and ensures comparability within the evaluation.

[24] In 2016, the differences in poverty levels between urban (13.9%) and rural areas (43.8%) was still significant.

Table 7: Sample characteristics of treated and control schools (2014)

	Treatment (n=4,389)		Control Group (n=25,002)		Mean Comparison
	Mean	**SD**	**Mean**	**SD**	**P-Value**
Output Variable					
Reading score	416.99	148.72	416.25	156.53	0.7720
Math score	498.14	72.74	498.42	69.72	0.8056
Control Variables					
Rurality	0.90	0.29	0.71	0.45	0.0000
Poverty	0.60	0.22	0.54	0.23	0.0000
Extreme poverty	0.30	0.19	0.24	0.18	0.0000
Malnutrition	40.26	15.96	35.63	15.69	0.0000
Multigrade school	0.56	0.49	0.39	0.49	0.0000

3.5 Design and results

3.5.1 Design

As described in Figure 10 and confirmed by the results of Table 7, the selection of treated schools follows a non-random selection process mostly based on poverty level, previous student achievement, and school type. This way of choosing schools generates a selection problem. This means that a bias is likely because both groups are not comparable in terms of observable characteristics that may have served to select the samples, which means that the difference in the outcome between treated and untreated may depend, instead of the treatment by itself, of such external characteristics. As argued by Vooren et al.

(2018), to solve this problem a quasi-experimental design such as propensity score matching (PSM) can be used to obtain unbiased results in evaluating teacher training programs.

PSM generates the control group by modeling the probability of participation in the program on the basis of the observed characteristics of the schools. Two assumptions should be made to ensure accurate estimation. The first assumption, conditional independence, implies that the program is based on observable characteristics. This assumption is fulfilled in this thesis because the selection of the treated schools depends mainly on observable characteristics, such as districts' poverty levels, past ECE results, and the rurality of the schools' location. The second assumption, common support, is made if the P-score of the treated schools have comparable observations in the propensity score distribution of the non-treated schools. (Rosenbaum and Rubin, 1983)

This study decides between the Nearest Neighbor Matching (NNM) and the Radius Caliper Matching (RCM) algorithms, preferring RCM because it imposes a tolerance level on the maximum propensity score distance, caliper, which is relevant when the distribution is very different between the two samples. The value of the caliper used in this research, as recommended by Smith and Todd (2005), is 0.01.

3.5.2 Results

Using the PSM technique and differentiating between multigrade and full-grade schools, it is possible to estimate the impact of the program. The steps of the analysis are as follows:

1. Estimating the P-score matching for multigrade and full-grade samples

To estimate the p-score matching of each school, a probit model is used:

$$\boldsymbol{D} = \beta_0 + \beta_1 rurality + \beta_2 extreme\ poverty + \beta_3\ malnutrition + \beta_4\ JUNTOS + \beta_5\ VRAEM,$$

where D is the variable that takes the value of 1 when the school is treated (teachers receive training) and 0 otherwise, β_0 is the constant parameter of the model, β_1 is the parameter of the effect of rurality on the treatment selection, β_2 is the poverty parameter, β_3 is the effect of malnutrition, β_4 is the parameter that controls the effect of the JUNTOS cash transfer program, and β_5 shows whether the school is located in the Apurímac, Ene, and Mantaro rivers (VRAEM) area[25].

Table 8 illustrates the results for Math for multigrade and full-grade samples. For multigrade schools, all control variables are statistically significant except for the rurality coefficient, which may be because most multigrade schools are rural, and so there is not sufficient variance in this variable. In the case of full-grade schools, all variables except JUNTOS are statistically significant. Rurality is statistically significant, which is expected because we find full-grade schools both in rural and urban areas of the country. For Reading comprehension results of the Probit estimation are similar both for multigrade and full-grade samples.

Table 8: Results of the Probit estimation for Math in full-grade and multigrade schools (2014)

	School management			
	Multigrade schools		Full-grade Schools	
Variable	**Coefficient**	**SE**	**Coefficient**	**SE**
Constant	-0.993***	0.045	-2.062***	0.058
Rurality	-0.0057	0.040	0.503***	0.061
Extreme poverty	0.3435***	0.072	0.416**	0.219
Malnutrition	0.2009***	0.063	0.978***	0.155
JUNTOS	0.1919***	0.028	-0.112	0.086
VRAEM	0.2941***	0.043	0.278**	0.127

*Significant at 10% level /**Significant at 5% level / ***Significant at 1% level

[25] VRAEM is a Peruvian territory involved with high levels of poverty, violence and illegal activities.

Taking into account the value of the coefficients, the probit model for both equations is as follows:

$$\boldsymbol{Multigrade} = -0.993 + 0.3435 extreme\ poverty + 0.2009\ malnutrition + 0.1919\ JUNTOS + 0.2941 VRAEM$$

$$\boldsymbol{Full - grade} = -2.062 + 0.503 rurality + 0.416 extreme\ poverty + 0.978 malnutrition + 0.278 VRAEM.$$

Both results show that the selection of treated schools is based on observable variables, thus fulfilling the first condition of the PSM model, conditional independence. About the common support, it is shown in detail in Table 9.

Table 9: Common Support of the Probit estimation for Math and Reading comprehension in full-grade and multigrade schools (2014)

	Common Support			
	Multigrade schools		Full-grade Schools	
Treatment Assignment	Math: On support	Reading: On support	Math: On support	Reading: On support
Untreated	16,092	16,094	6,486	6,487
Treated	4,052	4,052	337	337
Total	20,144	20,146	6,823	6,824

Applying both probit models obtained in Table 9 to the data, the number of observations in the treatment and control groups after matching using Radius Caliper Matching (0.01) is for multigrade schools around 4,052 treated and 16,094 control schools, and for full-grade 337 treated and 6,486, control schools. By this is possible to fulfill the second condition of the PSM model and allow us to continue with the matching process as will be detailed below.

2. Matching and estimating results

Results of the matching process after comparing the means of the treated and control matched samples observed in Table 10, reported in terms of effect size, show that the effect of the program for multigrade and full-grade schools is

positive and significant, both for Math and Reading Comprehension. The teachers in-service program raises student achievement in Reading Comprehension by 0.148 standard deviation for multigrade schools and 0.233 standard deviation for full-grade schools, both statistically significant at the 1% level. In Math effects are also positive and show an increase in learning outcomes by 0.083 standard deviation for multigrade schools and 0.137 standard deviation for full-grade schools, all significant at the 1% level. In any scenario Reading Comprehension always has better results than Math, which can be explained by the specific characteristics of the program: more focused on improving teachers' pedagogical skills (see Chapter 2). About differences in results between full-grade and multigrade school, where the first ones always show higher impacts, this can be explained by the homogeneity of students in full-grade schools. There, all students are of the same age and are positively influenced by the peer effect (Epple and Romano (2011), Duflo, Dupas, and Kremer (2009), and Kang (2007)). In multigrade schools, teachers divide their classroom time between different groups, reducing their ability to provide adequate instruction to each student.

Table 10: Estimation results and robustness on Reading and Math (2014)

	Multigrade schools		**Full-grade schools**	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
Reading score	0.148***	0.018	0.233***	0.066
Math score	0.083***	0.018	0.137***	0.051

Robustness Analysis	**Multigrade schools**		**Full-grade schools**	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
Reading score	0.156***	0.023	0.198***	0.055
Math score	0.075***	0.021	0.120***	0.049

*Significant at 10% level /**Significant at 5% level / ***Significant at 1% level
Multigrade sample (treated: 4,052; control group: 16,094) and full-grade sample (treated: 337; control group: 6,486)

Although smaller, these positive results support those obtained by similar interventions that report impact in terms of effect size in developing countries. Angrist and Lavy (2001) found that a training program in nonreligious Israeli schools increased students' test scores, both in Math and Reading Comprehension by 0.25 of a standard deviation on average. Yoon et al. (2008), in a group of 9 papers that met What Works Clearinghouse evidence standards, found that the average effect size for both subjects was 0.54 for teachers receiving in-service professional development programs. Interestingly, the results of the Peruvian program are greater than those obtained by Bruns et al. (2018) for the teachers in-service training program of the Brazilian state of Ceará, which shows an impact of 0.081 of a standard deviation in Mathematics and of 0.047 of a standard deviation in Portuguese for first year secondary students.

Finally, these results differ from those obtained in 2013 when the period prior to the change of focus of the program towards multigrade schools was analyzed. In that analysis, the Ministry of Economics and Finance observed a greater impact in Math than in Reading Comprehension (0.2 vs .0.1 standard deviation) probably generated because previously the program was focused on intervening in urban areas (where full-grade schools are more common).

3. Robustness analysis

To determine the potential variance in results, a robustness analysis of the estimates is conducted, dividing the sample in quintiles and dropping the first and last quintile from the sample. Table 10 show that results are very similar to those obtained in the previous analysis: 0.156 standard deviation for Reading Comprehension in multigrade schools and 0.198 standard deviation in full-grade schools; and for Math, 0.075 standard deviation for multigrade schools and 0.120 standard deviations for full-grade schools. All of these findings are significant at 1%, and so the positive impact for both subjects are confirmed.

3.6 Conclusion

For the period 2018-2020 the Peruvian government is expected to expand the in-service program to cover all rural schools. So far, the program has been implemented at 8,726 schools throughout the country, representing a 37% coverage of all targeted schools in 2016. However, an impact evaluation has yet to be carried out following the program's change of focus in 2012. Moreover, the literature review does not provide much evidence as there is a shortage of evaluations of similar programs on the impact of teachers' in-service training in rural areas. Therefore, this chapter has sought answers on the impact of the Peruvian teachers' in-service program on the educational performance of second graders in multigrade and full-grade schools.

From a scientific perspective, results will contribute to the literature on impact evaluations of professional development programs on student achievement, particularly those that are focused on rural schools.

Using a quasi-experimental propensity score matching design, results suggest that the program has a positive and significant impact on Math and Reading Comprehension. For Reading Comprehension, the effect size is 0.148 standard deviation for multigrade schools and 0.233 standard deviation for full-grade schools. For Math, results are positive, with an effect size of 0.083 standard deviation for multigrade schools and 0.137 standard deviation for full-grade schools. For both subjects, all results are statistically significant at the 1% level. The robustness analysis confirms these results. It is worth noticing that, regardless of the area, results are always higher for Reading Comprehension. The difference may be explained by the characteristics of the training, which is focused on improving teachers' pedagogical skills rather than on content-based training, something already suggested by the literature review in Chapter 2 when there is a combination of the type of intervention of the professional development and the subject.

This analysis method also has a number of limitations. Firstly, the matching model used does not control for omitted variables that may change over time and affect the treated and control groups differently. From the analysis of the

prioritization process the selected sample, it is observed that there are no omitted time variables.

Secondly, in using data available from the ECE, it was only possible to estimate the impact of the intervention in Math and Reading Comprehension on second graders. However, in 2015, the program intervened in all grades levels of elementary schools across various subjects, so it can also have an impact on other educational levels and subjects, which could differ from those of second grade students.

Finally, there is a potential risk of biasing the results if there are spill-over effects of the treated teachers on the untreated ones within the different treatment areas, what happens if the teachers can share the knowledge acquired in the training with other colleagues. Due to the magnitude of the intervention and the available data, this cannot be measured, but in any case this would lead to an underestimation of the effects, which supports the thesis of the positive impacts of the program

In the following chapters, this thesis investigates further the impact of the length of the program on student achievement. It is clear that the one-year interventions produce positive results, but the Peruvian program assigns 3 years of intervention to each teacher without a theoretical justification for this timeframe. So, these positive results are maintained when the program intervenes for the second and third year? Determining the impacts for the following years will allow us to know the optimal duration of the program. Additionally, it is important also to consider the differential impacts of the components of teacher training. Is it necessary to complete the workshops and update workshops to achieve optimal results? Addressing the questions of the optimal duration as well as the impacts for each component can help the Peruvian government to determine the optimal structure of the program, expected policy result considering the scarcity of resources available in a country like Peru.

Chapter 4: The sustainability of the impact of teachers' in-service training on student achievement

"What is defeat? Nothing but education. Nothing but the first step to something better."

—Wendell Phillips

This chapter is based on Carrillo, C., Maassen van den Brink, H., and Groot, W. (2017). The sustainability of the impact of the Peruvian teacher in-service training program on student achievement over time. TIER WP 17-20.

4.1 Introduction

To answer the second question of this research we have addressed the impact of the Peruvian in-service training program on students' achievements. The positive results, both in Math and Reading Comprehension, suggest that the intervention is effective. However, results of the previous chapter on both full-grade and multigrade schools were obtained using a static analysis of data from 2014. Because, according to the protocol for 2013–2015, the program is supposed to train teachers for three consecutive years for 169–224 training hours per year (Ministry of Education-PELA Annex 2, 2015), it is essential to determine if these effects change when teachers receive more than one year of training.

The impact evaluation literature has not focused on the optimal length of professional training programs. As stated by King and Behrman (2009) in the labor sector, it is necessary that research establishes for how long treatment groups must be exposed to a program in order to obtain benefits from the intervention.

This chapter's objective is to assess whether the positive impact of the Peruvian teachers' in-service program improves when teachers receive consecutive training for more than one year, both in Math and Reading Comprehension. From a local policy perspective, results may offer insights into the effectiveness of the program duration and in determining its optimal length, so if it is advisable to establish the three consecutive years as the time of permanence in it. Additionally, results will contribute to literature on the length of training programs, both in the educational and labor sectors, complementing papers by Garet et al. (2001), Yoon et al. (2007), Kluve (2006), Card et al. (2010), and McGuinness et al. (2014).

This chapter uses educational data from 2014–2016 from ECE, SIGMA, and the Ministry of Inclusion's Poverty Map to perform a propensity score matching (PSM) analysis with a difference-in-differences analysis to estimate how

incremental the impact of the program is when it last more than one year. This study expects to find that the program has a positive and durable effect on student achievement, while also gathering key data on optimal training length.

The chapter is structured as follows. Section 4.2 discusses the literature on training program duration, both on the labor and educational sectors. Section 4.3 describes the data and descriptive statistics. Section 4.4 presents the design and empirical results, including a robustness check. Section 4.5 provides concluding remarks and policy recommendations.

4.2 Literature

The objective of a training program, used as a policy instrument in many countries to increase employability and educational achievements, is to develop beneficiaries' skills and generate an impact on income, welfare, capabilities, or employment (McGuinness et al., 2014). However, a program's effectiveness is dependent on its training type and duration.

About the type, in the educational field the systematic literature review in Chapter 2 showed that content-based and intense programs had a more significant impact, which coincides with what was proposed by Yoon et al. (2008) and Clewell et al. (2004). In the field of labor studies, similar results were obtained by Kluve (2006), Card et al. (2010), Haelermans and Borghans (2011), and McGuinness et al. (2014), who argue that the combination of on-the-job programs focused on specific groups are more likely to have a positive impact on employment rates in comparison to large-scale unfocused programs.

In terms of program length, findings are diverse. For labor studies, Card et al. (2010), using a meta-analysis to evaluate a sample of 97 studies on employment training, found that medium-term and long-term training (2–3 years of post-program horizon) was more effective than short-term programs (one year of post-program horizon) on the employment status of beneficiaries. McGuinness et al. (2014) emphasizes that on-the-job training must be sustained over time

to have a positive impact on employment. Furthermore, Fitzenberger et al. (2006) found a positive relationship between duration and long-term employment effects in Germany. Greenberg et al. (2003) also found similar long-term effects on wages, especially among women, for government training programs.

In contrast, Kluve et al. (2007), in labor training in Germany, found that the duration of training programs had a positive impact on employment, but only in programs of less than three months in length. Similarly, Hujer, Thomsen, and Zeiss (2006) found that programs that lasted for more than 6 months had a negative impact on employment. Vooren et al. (2018), in another meta-analysis of 55 experimental and quasi-experimental studies published in the last 25 years, found that training programs that last up to 6 months have more positive long-term effects (more than a year after the start of the program).

The previous chapter has illustrated that a long-term, sustained intervention, such as the Peruvian teachers' in-service program, has a positive impact on student achievements, which confirms the findings of Clewell et al. (2004). Similar results can be observed in Garet et al. (2001), who found that two measures of duration, contact hours and time span, have a positive effect on teachers' knowledge and skills, and when using the logical training model for professional development presented in Figure 3.1, eventually on student achievement. Moreover, Yoon et al. (2007) describes that the American Institutes for Research (AIR), as part of the national evaluation of the Eisenhower Professional Development Program,[26] concluded that duration, including contact hours and extensions, is a key feature of professional development programs that improve teaching.

Although there are more data on the positive impact of program duration in the

[26] Professional development program running from 1985 in the USA aimed to support intensive teacher training in elementary and secondary schools. The program includes special interventions for minorities, women, individuals with disabilities, and economically disadvantaged students.

educational and labor sectors, there is no clear framework to establish the optimal duration of professional development programs. So far there are only general references about possible time thresholds as in Yoon et al. (2008), who states that studies with more than 30 hours of intervention show positive results but do not specify whether those results increase or decrease as training continues. Similarly, Supovitz and Turner (2000), using hierarchical linear models, found that after 80 hours of intense, subject-content training, there were changes in teachers' behavior (more inquiry-based teaching practices[27]) and after 160 hours on the environment of teachers' classroom. In the Peruvian teachers' in-service program, the training lasts between 169–224 training hours per year, more than the average of similar programs evaluated in Yoon et al. (2008) and other Latin American policies such as "Si Profe" and ¨Aprendes¨. Therefore, it is essential to determine whether the programs have long-term positive returns, if those returns increase or decrease over time and if there is a threshold for such performance.

4.3 Data and descriptive statistics

To develop the duration analysis for Reading Comprehension and Math results for second graders, this chapter uses the information sources detailed in Chapter 3:

1. Census student assessment (ECE) data, which provides second graders' "Rasch" results together with data on other schools' characteristics;

2. The information and management system for learning improvement (SIGMA), which provides information on schools treated by the program. This data includes the compliance level of coaches' visits for every year the school was treated, a required information to determine how long the teacher was

[27] Refers to a teaching model based on research, questioning, and experimentation.

exposed to the program; and

3. The Ministry of Inclusion's Poverty Map, which provides socioeconomic data of poverty, rurality, and malnutrition at a district level.

To perform the duration analysis, it is necessary to manage the data and create a dataset that allows for the identification of schools treated for a one-year, two-year, or three-year period in order to compare them and determine the incremental effects of additional periods of training, both in multigrade and full-grade schools. These findings are shown in Figure 11. Using data similar to that of Chapter 3, but from 2014 to 2016, this study was able to obtain information from SIGMA for three consecutive years on multigrade schools. 4,052 schools were treated for one year, 1,544 were treated two consecutive years, and 688 were treated for three consecutive years. In the same period for full-grade schools, 337 schools were treated for one year, 13 were treated for two years, and 7 were treated for three years. There are little data on full-grade treated schools because the in-service program has not consistently intervened in these schools since the policy adjustment in 2012. Figure 12 is a comparison between treated samples from multigrade and full-grade schools. As observed, the scarcity of data for two-year and three-year training in full-grade schools must be considered in the development of the analysis.

Figure 11: Dataset construction process for the duration analysis

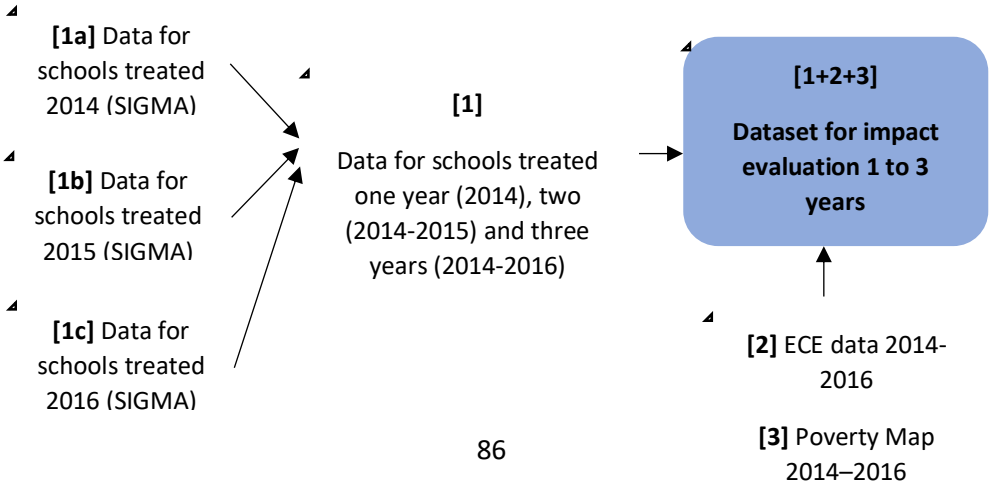
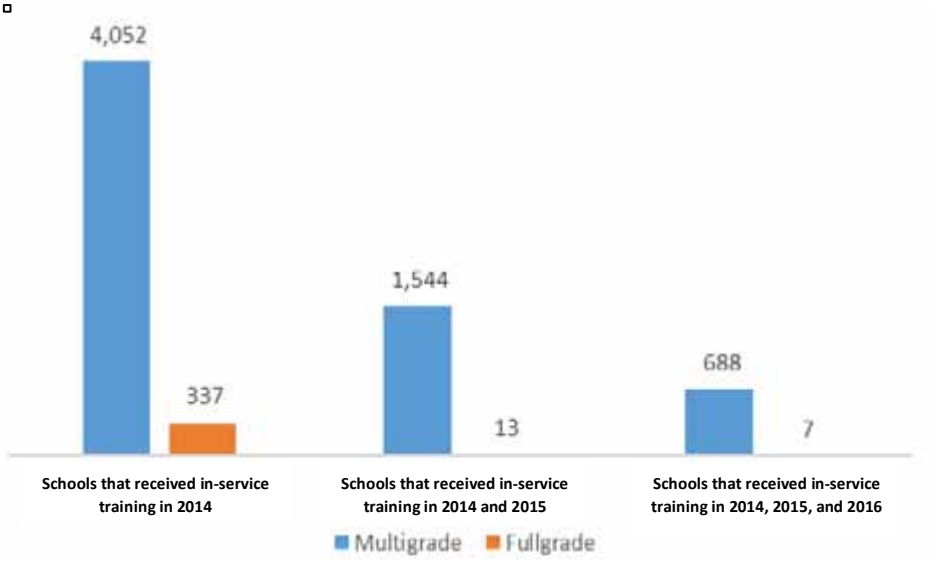


Figure 12: Schools who received in-service training for 1, 2, or 3 years for multigrade and full-grade schools



4.4 Design and Results

4.4.1 Design

In Chapter 3, the selection process for treated schools was specified as a non-random selection process based on poverty level, previous student achievement, and multigrade schools located in rural areas. It is therefore necessary to use a quasi-experimental research design. This analysis uses matching techniques and combines them with a difference-in-differences (DiD) estimator to identify the effect on students' results of one or two additional years of training in the treated multigrade and full-grade schools. As stated by Vooren et al. (2018), this method should address the selection problem in the evaluation of teachers' training programs. This analysis also requires a robustness check to cover problems that may arise from sample selection and the unobserved heterogeneity of the sample.

In the difference-in-differences analysis, D is the binary indicator that equals 1 if a school has received the teacher in-service program and 0 otherwise. Also, Y_t^1 denotes the value of the students' achievement if the school has received the program, and Y_t^0 if not. Then, given a set of observable economic and social attributes from the school (X), the average effect (ATT[28]) of being a treated-on Y is

$$E[Y_t^1 - Y_0^1 \mid X, D = 1] - E[Y_t^0 - Y_0^0 \mid X, D = 1].$$

Because $E[Y_t^0 - Y_0^0 \mid X, D = 1]$ is not observable and cannot be estimated accurately, the approximation of this term should be achieved using a matching method that can aid in comparing treated groups with non-treated groups and controlling for differences in the distributions of X between the groups (Angrist and Pischke, 2009).

In propensity score matching, similarly to Chapter 3, this chapter uses the Rosenbaum and Rubin (1983) technique to match observations on the basis of a single index that summarizes all data from covariates. This propensity score, $p(X)$, is the probability of treatment conditionals on observable characteristics.

Therefore, the ATT can be computed as

$$E[Y_t^1 - Y_0^1 \mid p(X), D = 1] - E[Y_t^0 - Y_0^0 \mid p(X), D = 1],$$

where the non-comparability bias can be eliminated only by considering observations within the common support, the intersection of the distributions $p(X)| D = 1$ and $p(X)| D = 0$. As was performed in Chapter 3, for the identification of the ATT, this chapter chooses a set of conditioning variables X that are not influenced by the treatment: socioeconomic variables such as poverty, rurality, and exposure to social programs obtained from the ECE, SIGMA, and the Ministry of Inclusion's Poverty Map.

[28] Average Treatment on the Treated.

4.4.2 Results

Using the difference-in-differences analysis together with the propensity score matching (PSM), it is possible to estimate the impact of the program across multigrade and full-grade schools. The steps of the analysis are as follows:

1. Estimating the effect from one to three years of training using the PSM methodology.

As in the previous chapter, this chapter estimates the p-score matching of each school with a Probit model:

$$D = \beta_0 + \beta_1 rurality + \beta_2 extreme\ poverty + \beta_3\ malnutrition + \beta_4\ JUNTOS + \beta_5\ VRAEM$$

where D is the dependent variable that takes the value of 1 when the school is treated and 0 otherwise, β_0 is the constant parameter of the model, and β_{1-5} are the control parameters of the probit model (rurality, poverty, malnutrition, JUNTOS cash transfer program, and VRAEM location). The results, similarly to those in Chapter 3, show that the probit model depends mainly on those observable characteristics.

As observed in Table 11, results for one year of training[29] (2014) using Radius Caliper (0.01) and expressed in standard deviations show that the effect of the program on multigrade and full-grade schools is positive and statistically significant across both Reading Comprehension and Math.

Using the same methodology as that used in the previous stage, though taking into account the data on 2-year interventions (schools treated during 2014 and 2015) versus untreated schools[30], results are presented also in Table 11. The effect of the program on multigrade and full-grade schools in terms of standard

[29] With a common support sample of 4,052 for the treated and 16,092 for the untreated multigrade schools; and 337 schools treated and 5,486 untreated for full-grade schools.

[30] With a common support sample of 1,544 for the treated and 16,098 for the untreated multigrade schools; and 13 schools treated and 5,680 untreated for full-grade schools.

deviation is positive, significant, and larger than the effect of the one-year intervention, both for Math and Reading Comprehension.

In the case of 3-year interventions (schools treated from 2014 to 2016) versus untreated schools[31], the results in Table 11 show that the effect of the program on multigrade and full-grade schools is positive but insignificant in Math and Reading Comprehension. These results contrast with the ones obtained for one and 2-year analysis and provide evidence for the decreasing effect of in-service training after the second year of implementation. It is worth noting the reduced value of the betas, which suggests that the third year of the intervention has no effect on student achievement. This finding may be due to the reduced sample treated for three years (7 full-grade schools and 688 multigrade schools), which may have affected the results obtained, so results are examined further in the robustness analysis of the three-year intervention.

Table 11: Results of the PSM estimation from one to three years of training in multigrade and full-grade schools (2014–2016)

	Multigrade schools		**Full grade schools**	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
1 year (2014)				
Reading score	0.148**	0.018	0.233***	0.066
Math score	0.083**	0.018	0.137***	0.051
2 years (2014/15)				
Reading score	0.192***	0.016	0.291***	0.012
Math score	0.121***	0.017	0.148**	0.023

[31] With a common support sample of 688 for the treated and 10,098 for the untreated multigrade schools; and 7 schools treated and 5,680 untreated for full-grade schools.

	Multigrade schools		Full grade schools	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
3 years (2014/16)				
Reading score	0.025	0.096	0.095	0.099
Math score	0.020	0.098	0.027	0.093

*Significant at 10% level /**Significant at 5% level / ***Significant at 1% level

2. Estimating the marginal effect of 2- and 3-year training programs using a difference-in-differences (DiD) analysis.

To determine the marginal effect of the second year of training, a difference-in-differences analysis is used to compare results previously obtained with PSM 2014 and PSM 2014–2015. Similarly, results from the third year of training consider the subtraction of PSM 2014–2016 against PSM 2015.

As observed in Table 12, the two-year training has positive marginal effects, and its coefficients suggest the following:

- One additional year of in-service training has a positive and significant effect on student achievement in Reading Comprehension by 0.044 standard deviation in multigrade schools and 0.058 standard deviation in full-grade schools, both of which are statistically significant at the 1% level. In Math, the effect is also positive and shows an increase in learning outcomes by 0.038 standard deviation (significant at the 1% level) for multigrade schools and 0.111 standard deviation (significant at the 5% level) for full-grade schools. In both subjects, these incremental impacts are smaller than the results for the first year of training, which means that the additional years of training offer decreasing returns.

- Similar to the first year of training, Reading Comprehension has a larger marginal effect than Math regardless of school type.

For the incremental analysis of the third year of training, the results in Table 12 show that the third year does not generate a positive or significant impact on student performance, neither for Math nor Reading Comprehension, regardless of school type. In addition, the third year of intervention has a negative (insignificant) effect on the performance of second-grade students, which means that the investments made in these trainings are innocuous in enhancing student achievement.

The results obtained from the incremental analysis of the second and third years of the program are further examined by a robustness analysis.

Table 12: Results of the DD estimation: second and third additional years of training

	Multigrade schools		Full grade schools	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
2nd year (2014-2015)				
Reading score	0.044***	0.019	0.058***	0.025
Math score	0.038***	0.016	0.011**	0.030
3rd year (2015-2016)				
Reading score	- 0.167	0.098	- 0.196	0.095
Math score	- 0.101	0.099	- 0.121	0.090

*Significant at 10% level /**Significant at 5% level / ***Significant at 1% level

3. Robustness analysis

To assess the consistency of the results, a robustness check of the estimates is done, dropping the extreme values of the sample (first and last quintile). Table 13 shows the results of a difference-in-differences (DiD) analysis for the additional second year (2014–2015) and third year (2014–2016) of training, respectively. The results are similar and confirm the results obtained previously: positive and decreasing impacts on Math and Reading Comprehension for the second-year of the program and null results for the third year of intervention.

Table 13: Robustness analysis: 2 and 3 years of intervention (2014–2016)

	Multigrade schools		Full grade schools	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
2nd year (2014-2015)				
Reading score	0.039***	0.016	0.048***	0.021
Math score	0.035***	0.013	0.009***	0.049
3rd year (2015-2016)				
Reading score	- 0.126	0.097	- 0.159	0.095
Math score	- 0.079	0.098	- 0.116	0.099

*Significant at 10% level /**Significant at 5% / ***Significant at 1%

Finally, as there may be controversy about the negative effects of the third year of the intervention, which could have been caused by the small sample of schools (7 full grade and 688 multigrade), we return to do a full analysis, but this time focusing only on these schools, to observe the trend of these 3 years.

Table 14: Effect of training on schools with 3 years of intervention (2014–2016)

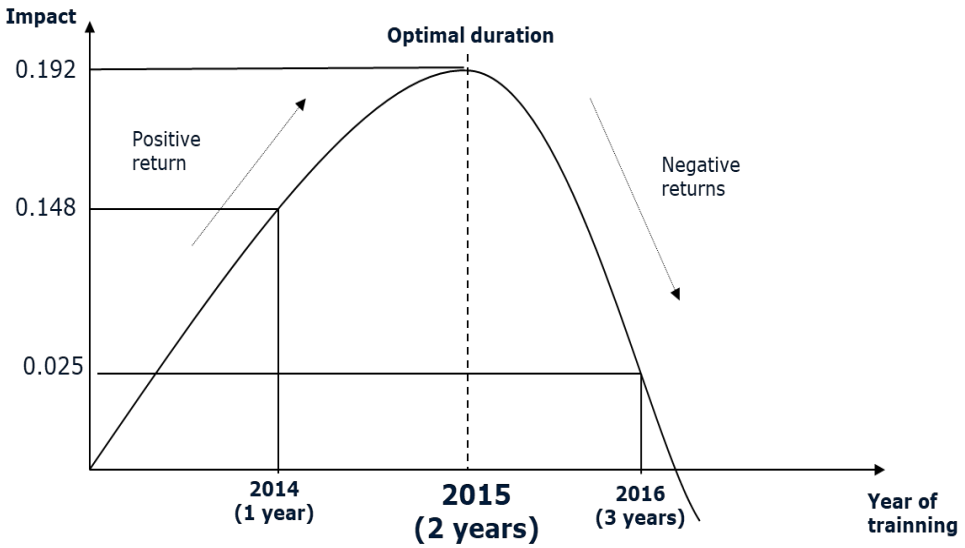
	Multigrade schools		Full-grade schools	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
1 year (2014)				
Reading score	0.013	0.080	0.018	0.085
Math score	0.011	0.081	0.014	0.090
2 years (2014–2015)				
Reading score	0.030	0.089	0.097	0.096
Math score	0.021	0.088	0.029	0.091
DiD 2nd year				
Reading score	0.017	0.096	0.079	0.097
Math score	0.010	0.098	0.015	0.091
3 years (2014–2016)				
Reading score	0.025	0.096	0.095	0.099
Math score	0.020	0.093	0.027	0.098
DiD 3rd year				
Reading score	-0.005	0.097	- 0.002	0.098
Math score	-0.001	0.095	- 0.002	0.097

*Significant at 10% level /**Significant at 5% level / ***Significant at 1% level

As shown in Table 14, the effect of one additional year of training on the schools treated during the 3-year period (DiD – 2nd year) is positive but not significant, most likely due to the small sample size. On the contrary, the effect of the third year of the intervention (DiD - 3rd year) is negative and not significant, with an almost null coefficient. The relevant aspect of both results is that the previous trend in the analysis is maintained: an additional year of training shows decreasing positive impacts while the third year has zero impact. These results confirm what was obtained in the previous analysis: the optimal level appears to be two years of in-service training.

Considering these findings, it is possible to conclude that a two-year exposure to the program still has positive (albeit diminishing) and significant effects in both subjects and both types of schools. In contrast, a three-year exposure to the program shows null effects on students' achievement. Thus, the optimal duration of the program, as shown in Figure 13, seems to be two years.

Figure 13: Optimal duration of the Peruvian teachers' in-service program



4.5 Concluding remarks and policy recommendations

To evaluate the impact of a long-term intervention and to determine the optimal duration of the Peruvian teachers' in-service program were the main objectives of this chapter. It has been found that that a two-year training has a positive, but decreasing, marginal impact on second-grade students' achievements both in Math and Reading Comprehension in both full-grade and multigrade schools. These positive impacts disappear when the program runs over three years of exposure, so the optimal length of the program must be two years of intervention. These results are similar to previous studies that found

that programs sustained over longer periods (usually more than a year) show better results, although they have not defined a threshold over which the effects decrease. (Garet et al., 2001; Yoon et al., 2007; Card et al., 2010; McGuinness et al., 2014; Fitzenberger et al., 2006).

The policy recommendation drawn from this result is that for Peru, this chapter finds that the optimal length of its long-term in-service program for coaching of teachers is two years. This threshold will allow the government to plan this educational intervention in addition to making better use of the available resources in order to cover part of the potential population of schools (63% of the total) not yet covered by the program.

Finally, now that the thesis has found the optimal length of the program, it is important in the next chapter to consider the completion rates of the different steps of the training program (i.e., the classroom visits, micro workshops, and update workshops) in order to determine which components of the training are key to achieving positive results. This analysis will give policy makers more insight into how to increase the efficiency of the allocation of public funds in the program.

Chapter 5: Program composition and its effects on student performance

"Good teaching is one-fourth preparation and three-fourths theater."
—Gail Goldwin

This chapter is based on Carrillo, C., Maassen van den Brink, H., and Groot, W. (2018). The effectiveness of different components of a teacher in-service training program in Peru. TIER WORKING PAPER SERIES. TIER WP 18-04

5.1 Introduction

This study has found not only that the Peruvian teachers' in-service program positively affects student achievement, but also that these positive impacts remain when the program is sustained over time up to an optimal length of two years. These results add to the evidence generated by studies like Bando and Li (2014), Bacolod and Tobias (2006), and Glewwe et al. (2004) on the impact of in-service programs in urban and rural environments in developing economies. These results also provide a basis to rely on this form of training to improve learning outcomes in multigrade and full-grade schools specially on areas of high income inequalities.

In the introductory chapter, we have described the Peruvian teachers' in-service program as focused on strengthening pedagogical competences rather than specific content-focused skills. This focus is the reason why the program's effects (Chapter 3) are greater in Reading Comprehension than in Math, a result that coincides with the findings of the systematic review of the literature of Chapter 2. But, even beyond the Peruvian program's tendency towards pedagogical training, there are differences among the program's components. On the one hand, it has classroom visits and micro workshops more focused on direct contact between the teachers and the coach with a "critical perspective." On the other hand, it has update workshops, aimed at helping teachers learn better pedagogical techniques, that apply a "traditional" teaching methodology more related to group meetings. These components of the program can generate differentiated impacts that are important to discover so that the government can optimize the intervention if necessary. About the differentiated impacts of the training components, the available literature has mainly focused on the overall impact of the intervention and has given less attention to the differential effects of the characteristics of the program (McGuinness et al., 2014). Thus, the fourth question of this research about the homogeneity of the contribution of these programs' components is still unanswered; some additional work is needed in order to respond to the question.

This chapter examines separately each of the Peruvian teachers' in-service program components to determine if the contributions of each component to the positive impact of the program are homogeneous or if it is possible to desist from any particular component while still achieving the same positive effect on students. This examination will allow us to propose an optimal design for the program. As in previous chapters, this analysis uses ECE, SIGMA, and the Ministry of Inclusion's Poverty Map data for 2015 to perform a difference-in-differences (DiD) analysis with propensity score matching (PSM) to estimate the differential impacts of each program component.

This chapter proceeds as follows. Section 5.2 describes the literature about the relation between the characteristics of training programs and their effects on student achievement. Section 5.3 describes the data and the descriptive statistics. Section 5.4 presents the results, giving special attention to the threshold of the "completion rate" for the program components. Moreover, it includes a robustness check of the estimates. Here also an optimal design of the program is developed considering the cost of every component of the training. Section 5.5 concludes and discusses some policy recommendations.

5.2 Literature

As stated before, professional development programs are systematic interventions aimed to improve teachers' knowledge. Darling-Hammond and McLaughlin (2011) recommends that this training needs to be based on concrete tasks and the discussion of teachers' experiences with students. Also, Supovitz and Turner (2000) reports that the development of teachers conducted in isolation from the classroom has a low impact on student performance and Lieberman (1995) describes that professional development should include discussion among colleagues within the school to get positive impacts. These recommended characteristics for the intervention are similar to those included in the methodology of the Peruvian intervention, that takes the form of personal coaching, workshops and school-based activities. The most important one, personal coaching – the basis of the Peruvian in-service program – can be

included into the definition of “on-the-job training,” an intervention designed to improve beneficiaries’ performance by supporting them in their daily routine, wherein someone who knows how to do the specific assignment shows the beneficiary how to perform it (Robbins, De Cenzo, and Coulter, 2016). This kind of methodology differs from off-the-job training, where the training is given at a site away from the work environment and uses lectures, case studies, simulation, role playing, etc.

Regarding on-the-job training, a number of studies have found that this type of training has positive effects mainly in the labor field. Haelermans and Borghans (2011), in a meta-analysis of studies published between 1981 and 2010, found a positive impact on salaries of up to 2.6% for each training received. Card et al. (2010), in a meta-analysis of 97 studies, found that on-the-job training programs appear to have more positive results in the medium run than in the short run on employment rate. Bartel (2015) also found positive impacts, specifically on the wages and performance (productivity) of a manufacturing firm, after controlling for selection bias in the assignment of the training. De Grip and Sauermann (2011), using an experimental design that randomly selects the workers of a Dutch telephone company for receiving an on-the-job training during 10 half-day sessions by a coach, found a 9% impact in workers productivity. Other authors like Fitzenberger and Völter (2007), Conti (2005), Dearden et al. (2006), and Mason et al. (2012) have confirmed that enhancing professional skills through on-the-job training produces positive results and leads to sustained employment, labor productivity, and wage effects. Those findings are robust to several estimation strategies and modeling specifications.

In the educational field, several in-service interventions have been previously considered: Angrist and Lavy (2001) considered coaching sessions; Swinton, Scafidi and Woodard (2012) considered workshops; and Bando and Li (2014) considered intensive courses. In all three cases, the in-service interventions showed positive impacts of on-the-job training on student achievement. Meta-analyses like Yoon et al. (2008) have also confirmed the results of these kinds of interventions.

For this thesis, it is important to focus on the characteristics of on-the-job

training and how the contact between the coach and the teacher supports the teacher's daily routine, unlike the distant relationships that exist in off-the-job training. As can be observed in Table 6 in Chapter 3, not all the training activities used by the Peruvian training program fall within the definition of on-the-job training, mainly because of the relation and nature of the interaction between the teachers and coaches within each program component. According to the protocol for the period 2013–2015 summarized in Table 6, for classroom visits, the coach observes the teacher practicing during lecture hours. The coach's intention is to give critical feedback to the teacher afterwards to improve the teacher's skills; this interaction thus constitutes a direct coach–teacher contact. For the micro workshops, the monthly meetings between the coach and his group of teachers are intended to enable discussion about the teachers' classroom experiences and the situations that occurred within the lecture hours; the micro workshops thus also constitute direct coach–teacher contact. These two components are close to the definition of on-the-job training. In contrast, the update workshops take a different approach. These meetings, which are usually located outside the area of influence of the school and are given by the trainer (who doesn´t have any previous contact with teachers), have the objective of training teachers on different pedagogical topics of interest. Most of the time, the update workshops are not specifically linked to the teachers' reality. Hence the update workshops are closer to the definition of off-the-job training.

Accordingly, the following sections assess the differences in student impact between the components that have a direct relationship between the coach and teacher vs. the component with an approach more distant from the day-to-day routine in the classroom.

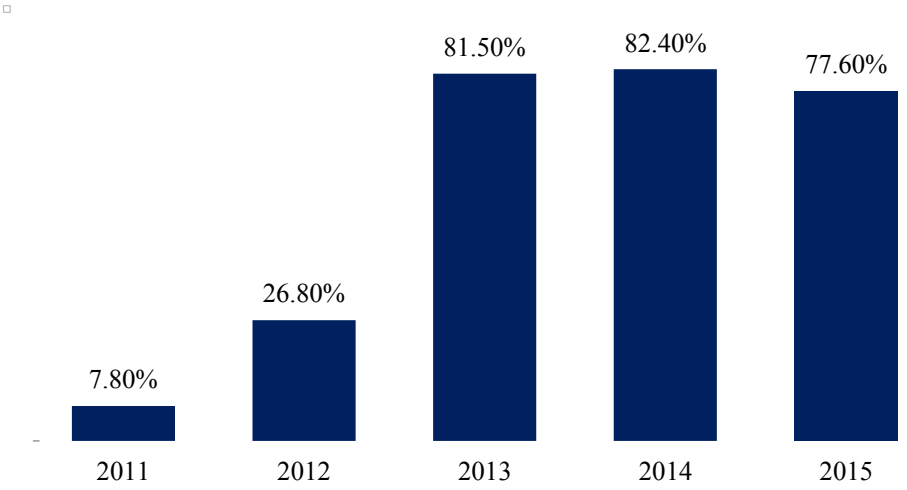
5.3 Data and descriptive statistics

The data used in this chapter are taken from three main sources: ECE, SIGMA, and the Ministry of Social Inclusion's Poverty Map for a single year, namely 2015. As the objective of this chapter is to assess the effect of the different program components (i.e., classroom visits, micro workshops, and update workshops) on student achievement, it is important to first analyze the

completion rate of the program. Figure 14 shows that the number of teachers who complete the training has been improving since the beginning of the intervention, rising from an initial 7.8% in 2011 up to 77.6% in 2015[32]. In terms of attended schools and trained teachers, this means that 11,908 schools and 25,786 teachers received training in 2015 (Ministry of Education, 2016).

Given that we have a sufficient number of trained teachers, it is necessary to define a threshold for the treated and non-treated samples for each component. According to SIGMA 2015, 97% of intervened teachers complete more than 50% of all the components, so this is the threshold used to consider a "complied" component for the analysis.

Figure 14: Completion rate of the in-service training for teachers 2011–2015



Source: Ministry of Education, 2016.

Table 15 presents the sample sizes for the treated and non-treated teachers using the threshold of 50% completion for each component. The presentation of the results differentiates the level of compliance per subgroup. Treated

[32] For multigrade/full grade, completion rates for each component in 2015 were as follows: classroom visits: 78%/89%; micro workshops: 75%/70%; update workshops: 77%/80% (Ministry of Education, 2016).

teachers are broken down into four groups: G1, or those that completed all the components with more than 50% attendance; G2, or those who completed classroom visits and micro workshops with more than 50% attendance; G3, or those that completed classroom visits and update workshops with more than 50% attendance; and G4, or those that completed micro workshops and update workshops with more than 50% attendance. The non-treated teachers group includes teachers who did not complete any of the components with more than 50% attendance and those teachers who did not receive any kind of intervention.

Table 15: Sample sizes of treated and non-treated teachers (2015)

	Multigrade	**Full grade**
Not treated (< 50% completion in all)	16,700	6,601
Treated (> 50% completion)		
G1: All	7,207	1,214
G2: Visits–Micro	158	26
G3: Visits–Update	267	13
G4: Micro–Update	88	10
Total	**24,420**	**7,864**

5.4 Results

This chapter uses the same two-step PSM analysis proposed by Rosenbaum and Rubin (1983) as used in previous chapters to address the selection problem in the evaluation of teachers' training components. The chapter also contains a robustness check to cover any problems that may arise from the heterogeneity of the sample.

1. Estimating the effect of training components for multigrade and full-grade schools using propensity score matching (PSM).

We estimate the p-score matching of each school with a probit model, differentiating the sample in multigrade and full-grade schools.

$$D = \beta_0 + \beta_1 rurality + \beta_2 extreme\ poverty + \beta_3\ malnutrition + \beta_4\ JUNTOS + \beta_5\ VRAEM,$$

where D is the selection variable that takes the value of 1 when the school is treated (teachers receive the training) and 0 in the other case; β_0 is the intercept of the model; β_1 is the parameter of the effect of rurality on the treatment selection; β_2 is the poverty parameter; β_3 is the effect of malnutrition; β_4 is the parameter that controls the effect of the JUNTOS cash transfer program; and β_5 shows if the school is located in the VRAEM vulnerable area. Results of the probit model depend on these observable variables.

As observed in Table 16, using Radius Caliper (0.01) and data for 2015, results for the complete intervention (G1) indicate positive and significant impacts: The effect sizes for Reading Comprehension are 0.179 for multigrade and 0.277 for full-grade schools, both statistically significant at 5% and 10% level respectively. For Math the results show the same pattern: an effect size of 0.117 for multigrade and 0.133 for full-grade schools, both statistically significant at 5% and 10% level respectively. These results are congruent with the ones obtained in Chapter 3 for a one-year complete intervention. For the other partial combinations of program components, namely G2, G3, and G4, Table 16 shows the values of the effect sizes on student performance. It is noteworthy that the combination G2 (completed classroom visits + micro workshops) shows results very similar to the complete intervention (G1), which advances the importance of having classroom visits and micro workshops if we want to have positive and relevant results for the intervention.

Focusing on the impact of each component, Table 17 shows that for Reading Comprehension, the single effect size of the classroom visit component is 0.078

for multigrade and 0.119 for full-grade schools, both significant at 10% level[33]. For the micro workshop, the single effect size is 0.069 for multigrade and 0.104 for full-grade schools, both significant at 10% level, while for the update workshops, the single effect size is 0.004 for multigrade and 0.092 for full-grade schools, both significant at 10% level. Results for Math show a similar pattern on a smaller scale since the impact of the intervention on Math is smaller. This is consistent with the findings in Chapter 3.

Table 16: Results of the PSM estimation for Reading Comprehension and Math during 2015 in multigrade and full-grade schools.

	Multigrade		Full grade	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
G1: Visits–Micro–Update				
Reading score	0.179**	0.015	0.277*	0.022
Math score	0.117**	0.009	0.133*	0.018
G2: Visits–Micro				
Reading score	0.175**	0.017	0.185*	0.072
Math score	0.109**	0.016	0.134	0.068
G3: Visits–Update				
Reading score	0.110**	0.014	0.173*	0.091
Math score	0.105*	0.010	0.127	0.090
G4: Micro–Update				
Reading score	0.101*	0.025	0.158*	0.098
Math score	0.100*	0.023	0.122	0.097

*Significant at 10% level / **Significant at 5% level /***Significant at 1% level

[33] As G1 is the full package (classroom visits+ micro workshops+ update workshops), and G4 is composed of micro workshops + update workshops, the individual effect size of classroom visits in Reading Comprehension can be obtained by subtracting G1 - G4 = 0.179 - 0.101 = 0.078.

Table 17: Results of the impact of every component of the training in multigrade and full-grade schools.

	Multigrade		Full grade	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
Classroom Visits				
Reading score	0.078*	0.025	0.119*	0.021
Math score	0.017*	0.021	0.011*	0.024
Micro workshop				
Reading score	0.069*	0.021	0.104*	0.062
Math score	0.012*	0.026	0.006	0.069
Update workshop				
Reading score	0.004*	0.025	0.092 *	0.025
Math score	0.008*	0.023	-0.001	0.090

*Significant at 10% level / **Significant at 5% level /***Significant at 1% level

The results after the matching show that for multigrade and full-grade schools, "classroom visits" is the component with the greatest impact on both Math and Reading Comprehension, followed closely by the micro workshops. For the component of update workshops, even though its impact is still positive and statistically significant (with the exception of the impact of full-grade schools on Math, with null effects), it is clearly the least important of the three components for second graders.

These different results for the training components can be explained by the characteristics of each component in terms of the strategy of the intervention. As observed in Table 6 in Chapter 3 and as described above, classroom visits and micro workshops use direct interaction between the coach and the teachers and are usually carried out within the school or in the direct neighborhood, so these components offer the possibility to provide feedback from a critical perspective, similar to on-the-job training. In contrast, the update workshops require the mobilization of teachers, putting teachers into a different environment and

using a technique more related to off-the-job training. This finding is thus consistent with Bartel (2015), Mason et al. (2012), Angrist and Lavy (2001), and Bando and Li (2014), who have all highlighted the positive impacts of on-the-job training in the fields of labor and education.

Finally, an additional but important result obtained from the analysis of Tables 5.2 and 5.3 is that the effects of having a combination of components such as classroom visits and micro workshops are greater than the sum of the individual effects for each component alone. Thus, the result of the previous combination in multigrade schools for Reading Comprehension (Table 16) is 0.175, significant at 5% level, which is larger than the sum of both individual effects observed in Table 17, which is 0.147. Similar results are observed for the other combinations (visits + update workshops, micro workshops + update workshops), suggesting the generation of multiplicative effects when more than one component of the training is combined.

2. Robustness check

As the treated sample was defined by a threshold of more than 50% compliance, it is important to analyze the sensitivity of the results to the choice of the threshold. We do this by estimating again the parameters using different thresholds. Figure 15 presents the sample size for each group of components by different levels of the threshold. As observed, when the threshold is less than 50%, the sample size of the treated groups remains almost the same. In contrast, when the threshold is greater than 80%, some of the treated groups (especially the complete intervention-G1-sample) are substantially smaller. This analysis indicates that a balanced sample size can be found between a threshold of 60% and 80%. The robustness analysis is conducted taking this finding into consideration.

Figure 15: Results of the robustness analysis of the PSM estimations

Table 18 presents for Reading Comprehension the results considering three different thresholds (60%, 70%, and 80%). The estimates are very similar to the ones obtained in the previous analysis, both in values of the coefficients and in levels of significance. For Math, we find the same. As the results remain similar, it can be concluded that the results obtained earlier appear to be robust.

Table 18: Robustness analysis results of the PSM estimation (Reading Comprehension)

	Multigrade		Full grade	
Output	**Coefficient**	**SE**	**Coefficient**	**SE**
Reading				
G1: Visit–Micro–Update				
60%	0.181**	0.016	0.277*	0.022
70%	0.179**	0.007	0.276*	0.018
80%	0.177*	0025	0.274*	0.022
G2: Visits–Micro				
60%	0.179*	0.022	0.184*	0.070
70%	0.178*	0.024	0.182*	0.067
80%	0.176*	0.026	0.179*	0.058
G3: Visits–Update				
60%	0.116*	0.023	0.173*	0.090
70%	0.117*	0.024	0.173*	0.087
80%	0.118*	0.021	0.172*	0.092
G4: Micro–Update				
60%	0.104*	0.025	0.157*	0.098
70%	0.104*	0.025	0.157*	0.098
80%	0.105	0.023	0.156	0.097

*Significant at 10% level / **Significant at 5% level /***Significant at 1% level

The conclusion is that the PSM with DiD analysis for each component shows that the classroom visits and micro workshops have a greater impact than the update workshops. As stated before, this result is most probably due to the type and nature of each training component and the intensity of the interaction between the coaches and teachers in the classroom visits and micro workshops.

3. Optimal design

The objective of this chapter also includes recommending the optimal design for the program. Accordingly, a cost-effectiveness analysis is performed to establish the comparative effect of each of the program components and thereby form policy recommendations. Using official cost-reports of the Ministry of Education for treated multigrade schools,[34] Table 19 shows that update workshops is the most expensive component, both per student and per teacher, followed by classroom visits and micro workshops . On average, each program component's cost for one year of intervention is US$ 68 per student. These costs include travel costs, didactic material used in coaching, expenses for seminars and workshops, and other costs related to the visits and the coaching planned during the year.

Table 19: Costs of program components for multigrade schools (2015)

Activities	Total Costs (US$)	Cost per student (US$)	Cost per teacher (US$)
Classroom visits	12,506,142	71.83	869.0
Micro workshops	7,612,327	43.72	528.9
Update workshops	15,413,932	88.53	1,071.0

Source: Ministry of Education Official Reports, 2017.

Using this information together with the effect size of the grouped and individualized components in Table 17, it is possible to perform the cost-effectiveness analysis. Standardizing the cost per student per 0.1 standard deviation (SD), it is possible to obtain – in an individualized analysis – the information presented in Table 20. As observed in the table, and taking multigrade schools as an example, for Math, the cost per pupil to increase students' results by 0.1 SD is US$ 174.4 for the whole intervention, while for

[34] Information for full-grade schools was not available.

Reading Comprehension that cost is US$ 114.0.

Individualizing by components, for multigrade schools, the cost per pupil to increase students' results in Math by 0.1 SD is US$ 422.5 using classroom visits, US$ 364.4 using micro workshops, and US$ 1,106.6 using update workshop. For multigrade schools in Reading Comprehension, the results are US$ 92.1, US$ 63.4, and US$ 2,213.3 for 0.1 SD improvement obtained via classroom visits, micro workshops, and update workshops, respectively. These results demonstrate that the most cost-effective component is micro workshops, followed by classroom visits and update workshops, both for Math and Reading Comprehension. However, these results also show that the complete intervention, due to its multiplicative effects, has better or at least similar cost effectiveness.

To dig deeper into the question of optimal design, it is necessary to perform an additional analysis. This time, though, components are grouped because of the multiplicative effects of combining components within the intervention. Doing this analysis for Math, it is possible to find that the combination of classroom visits and micro workshops costs US$ 106.0 per student per 0.1 SD, making this combination the most cost-effective combination for that subject. For Reading Comprehension, the situation is similar: grouping classroom visits and micro workshops costs US$ 66.0 per student per 0.1 SD, making this combination cheaper than complete intervention (US$ 114.0 per student). Thus, the effectiveness of the intervention can be increased if the program is reconfigured to include only the first two components.

Table 20: Cost-effectiveness analysis for multigrade schools

Intervention	Effect Average test score (SD)	Cost Cost per student (dollars)	Cost-effectiveness ratio Cost per student per 0.1 SD gain (dollars)
Math			
Separate components			
Visits	0.017**	71.83	422.5
Micro	**0.012****	**43.72**	**364.4**
Update	0.008**	88.53	1,106.6
Group of components			
Visits–Micro–Update	0.117	204.08	174.4
Visits–Micro	**0.109**	**115.55**	**106.0**
Visits–Update	0.105	160.36	152.7
Micro–Update	0.100	132.25	132.3
Reading Comprehension			
Separate components			
Visits	0.078**	71.83	92.1
Micro	**0.069****	**43.72**	**63.4**
Update	0.004**	88.53	2,213.3
Group of components			
Visits–Micro–Update	0.179	204.08	114.0
Visits–Micro	**0.175**	**115.55**	**66.0**
Visits–Update	0.110	160.36	145.8
Micro–Update	0.101	132.25	130.9

*Significant at 10% level / **Significant at 5% level /***Significant at 1% level

5.5 Conclusion

The aim of this chapter was to analyze the effect of each of the components of the Peruvian teachers' in-service program on student achievement in order to discover if the design of the program is efficient. Using propensity score matching with difference-in-differences analysis, we found for both Math and Reading Comprehension a strong relation between students' achievement and the two components that strengthen the coach–teacher relationship and enable discussion of what happens inside the classroom, namely the classroom visits and micro workshops. Update workshops appear to have not a relevant effect. Additionally, we found that grouped components have larger effects than each component individually. With these results we get closer to what mentioned by Supovitz and Turner (2000) and Cohen and Ball (1990): that professional development (understood as a training intense, sustained and carried out close to the classroom environment) is the best option to change teaching practices, because the traditional policies and programs that regulate teaching behavior without greater contact with them, have no major impact on their behavior or daily routine.

By performing a cost-effectiveness analysis, we identified the optimal program design. The results show that a combination of classroom visits + micro workshops provides better results in terms of US$ per 0.1 SD gain. The intervention could therefore be improved if the program is reconfigured to include only those two components. With respect to update workshops, the analysis leads to the conclusion that, instead of investing in this component, it may be better to re-allocate these resources to other interventions aimed at improving students' performance.

Chapter 6: Conclusions and discussion

"Sine doctrina vita est quasi mortis imago."

(Without learning, life is but the image of death)
—Dionysius Cato

6.1 Introduction

This thesis puts forth two main reasons why the Peruvian government should invest in the improvement of education, particularly when the program under discussion focuses on areas of high inequalities. First, there is a positive relationship between skills acquired during primary and secondary education, future achievements, and a country's economic growth (Hanushek and Luque, 2003; Hanushek and Woessmann, 2012). Second, higher-quality education reduces educational inequalities, which in turn has an impact on future reductions of income inequalities (Blau and Kahn, 2005; Bedard and Ferrall, 2003; Checchi and van de Werfhorst, 2017). The Peruvian government, since 2008, has implemented the Strategic Program for ¨Learning improvement of regular basic education students¨ (PELA), the objective of which is to improve learning achievements in students from 3 to 16 years old by intervening in four areas: i) school infrastructure, ii) teacher development, iii) educational material, and iv) monitoring and evaluating of learning and educational quality. The second policy area, namely teacher development, includes an in-service teachers' professional development program. This program aims to improve the pedagogical practices of teachers working in the context of high income inequalities through a systematic process of meetings between a coach and teachers undergoing the intervention. The following elements are involved in this program (Ministry of Education, 2014):

- Classroom visits, wherein the coach observes teachers' performance in classrooms and identifies pedagogical skills that must be addressed, giving monthly feedback on teachers' practices;
- Micro workshops, which are monthly meetings where the coach and trained teachers share experiences and investigate different pedagogical topics;
- Update workshops, which are usually located outside the schools' area of influence, and where a large group of teachers meets with the objective of updating their knowledge on pedagogical topics.

The main objective of this thesis was to analyze the impact of the Peruvian teachers' in-service program on the performance of second-grade students. The results of this analysis will provide policy advice to the Peruvian government as a decision is made as to whether or not to expand the program's coverage to other vulnerable areas. This thesis also contributes to broadening the literature on the relationship between professional development programs and student achievement.

The available evidence on the impact of professional development programs on student achievement is incomplete. Research is outdated and frequently does not use a clear and rigorous quantitative methodology in identifying the relationship between professional development and student achievement. The last systematic review of evidence that included rigorous standards (inclusion criteria), such as the use of experimental or quasi-experimental studies, and that was published in recognized journals and used reliable databases for analysis was conducted 10 years ago (Yoon et al., 2008). Therefore, a new review of the literature is necessary, which through the use of rigorous inclusion criteria allows us to know the state of the art of the impact of teacher training on student performance.

If assess the impact of the program, we also have to link it to the scale of the resources used and to the targeting strategy employed to cover the majority of vulnerable schools. In terms of resources, the program is part of the PELA strategy, which has enacted the largest educational intervention ever in the Peruvian public sector with a budget expenditure of almost 58% of the US$ 8.3 billion allocated to the education sector in 2016. In terms of targeting, as of 2016, this program covered 8,726 full-grade, multigrade, and bilingual primary schools throughout the country, or 37% of the total number of vulnerable schools in Peru.

Within this thesis, Chapter 2 presented a systematic literature review of empirical research from the last 20 years on the impact of professional development programs on student achievement. The chapter identified specific characteristics that may increase the probability of the program's positive impact on student outcomes. Chapter 3, using propensity score matching (PSM),

assessed the impact of the Peruvian teachers' in-service program on student results. This chapter explored differentiated impacts on full-grade (urban) and multigrade (rural) schools, which will provide the government with information about the intervention's ability to improve student results in the context of high income inequalities. Chapters 4 and 5, combining the propensity score matching (PSM) and difference-in-differences (DiD) analysis, provided more in-depth insight into specific characteristics of the in-service program to determine whether it was possible to improve its designs and make it more efficient. Chapter 4 assessed whether the impact of the program could be sustained over time, while Chapter 5 assessed each of the program's components to determine if their contribution to student results was homogeneous. Results from those chapters defined an optimal training duration and the need for the three components of the program to have the expected impact. Both findings could facilitate government policy decisions and improve the efficiency of the program.

Section 6.2 of this concluding chapter uses the findings of previous chapters to draw conclusions. Section 6.3 analyzes the policy recommendations of the improvement of the program. Section 6.4 discusses the limitations of the study. Finally, section 6.5 proposes suggestions for future research.

6.2 Overall Conclusions

The introduction of the thesis presented four research questions. Table 21 summarizes the main contributions of the thesis, describing each research question and the main findings.

Table 21: Main contributions in terms of content per chapter

Research question (chapter)	Main findings
Is there systematic evidence of the impact of the government's professional development program of teachers on student achievements? (Chapter 2)	51% of the estimates obtained from impact studies of professional development programs have a positive and significant impact on student achievement. Some characteristics of the training increase the probability of a positive impact, such as focusing on improving Math rather than Reading Comprehension, using content-based programs rather than pedagogical ones, and an intensive program sustained for a longer period of time (+60 hours). Additionally, the literature review finds that Math can be improved with the use of content-based programs, whereas Reading Comprehension responds more to pedagogical programs. Characteristics like location (urban/rural) and development level of the country are not related to the impact on student performance.
What is the impact of this professional development program of teachers in Peru? (Chapter 3)	Impact evaluations show positive and significant (at the 1% level) impacts for Math and Reading Comprehension. For Reading Comprehension (effect size): 0.148 and 0.233 standard deviation for multigrade and full-grade schools respectively. For Math (effect size): 0.083 and 0.137 standard deviation for multigrade and full-grade schools respectively. The robustness analysis confirms the results.
What is the optimal duration of the program in order to maximize its positive impact on treatment groups? (Chapter 4)	The optimal duration is a two-year intervention. Using the PSM and DiD analyses, it is found that a two-year training program has a positive-but decreasing- marginal impact on second graders, both in Math and Reading Comprehension in full-grade and multigrade schools. In the third year, the positive impact disappears.
Is the contribution of the professional development program's components to the	The first two components, classroom visits and micro workshops, have a strong relationship with student achievement. Update workshops appear to have an

Research question (chapter)	Main findings
positive impact on Reading Comprehension and Math homogeneous? (Chapter 5)	almost null effect on student results. Reason seems to be the way the coach interacts with teachers. For the first two components there is a direct relation between coach and teachers, close to an on-the-job intervention. Update workshops use a traditional off-the-job contact. Additionally, the thesis found that grouped components had a larger impact than each of them individually.
Is the Peruvian in-service teachers' program cost efficient when compared to other educational interventions? (Chapter 6)	The Peruvian in-service training is more cost efficient than a comparable local intervention (JEC) and is among the more cost-efficient interventions when compared to other US public educational interventions. This program yielded better results than programs intended to increase teacher salary, voucher programs, and those intended to reduce class size. The program yielded results comparable to those of programs intended to emphasize the methodology of learning by doing.

The conclusions of this thesis are organized into six statements below. These statements combine insights from different chapters and provide overarching conclusions based on the findings of this paper.

1. The effect size of the Peruvian in-service teachers' program is smaller than the effect size of similar interventions found in the international literature.

The literature review in Chapter 2 on professional development programs established that 51% of the estimates taken from the interventions analyzed had a positive and significant impact, which confirmed the positive effects of in-service training on student achievement.

For those studies that show standardized impacts, the effects size range from 0.05 standard deviation (Swinton, Scafidi, and Woodard, 2012) to 0.39 standard deviation (McCutchen et al., 2002). Yoon et al. (2008), in a group of 9 papers

that meet the What Works Clearinghouse research standards, found an average effect size of 0.54 standard deviation. Considering those information, the international average impact of professional development programs is around 0.4 standard deviation. These effect sizes are greater than the impact of the Peruvian in-service program obtained in Chapter 3. For Reading Comprehension, an impact between 0.14 and 0.23 standard deviation was obtained for multigrade and full-grade schools, and for Math an impact between 0.08 and 0.13 standard deviation for multigrade and full-grade schools was obtained.

These differences may be caused by the focus of the Peruvian program, which is delivered principally to rural multigrade schools. As observed in Chapters 3 to 5, school type accounts for 55% to 65% of the differences in impact in Math and Reading Comprehension, and the result is always lower when the program is delivered to rural multigrade schools. This result is mainly because the students in multigrade school are of different ages but located in the same classroom, which hinders the teachers' effective time for students in similar cohorts, and reduces peer effects. A peer effect is the positive effect generated in the student through interactions with their peers (of the same age or similar characteristics). Epple and Romano (2011), Duflo et al. (2009), and Kang (2007) found that an improvement in peer outcomes can have a positive impact on the student that varies between 20% and 60% of improvements, an impact that is lost in multigrade schools.

2. Contrary to the ambiguous results of general labor training programs, teachers' professional development interventions have a positive and significant effect when the interventions are sustained for a longer period of time.

Labor training is similar to educational training in that both types of training seek to improve the skills and competences of the beneficiaries. The goal of labor training is to improve the ability of workers to perform technical tasks or to relocate workers in a shorter time in the labor market. The goal of teacher training is to improve teaching skills. To establish a framework of the impact of the duration of training programs on the capabilities of the beneficiaries, this

thesis uses literature on the labor sector. Results from labor training programs´ duration on employment status and wages are inconclusive. Systematic literature reviews, such as Card et al. (2010) and Vooren et al. (2018), and studies such as McGuinness et al. (2014), Fitzenberger et al. (2006), and Greenberg et al. (2003) have found that long-term training programs are more effective. In contrast, Kluve et al. (2007) and Hujer et al. (2006) found negative employment effects in long-term programs.

Contrary to this relationship, in the education sector, the literature review showed that long-term programs had a more positive effect than short-term interventions. 62% of estimates that assess the effect of professional development programs of over 60 hours found positive and significant results, whereas approximately 50% of estimates found positive results for programs with less than 60 hours of training. These findings support findings from authors such Kennedy (1998), Supovitz and Turner (2000), Garet et al. (2001), Clewell et al. (2004), Yoon et al. (2007), and Navarro and Verdisco (2010), who also established the positive impact of sustained training in the educational sector. Positive results in the analysis of the Peruvian in-service teachers' program (a sustained intervention) in Math and Reading Comprehension supported these results.

The positive impact of the Peruvian program is a result of the following characteristics:

- more exposure time than similar interventions: from 169 to 224 hours of training per year, depending on whether the school is a multigrade or full-grade school;

- three years of intervention per teacher based on Ministry of Education guidelines.

These results are based on changes in teaching behavior, which according to the causal model based on Yoon et al. (2008), has a significant impact on students. This impact can be observed in what Supovitz and Turner (2000) describe as a

"culture of investigation", which happens when the teacher encourages discussion and collaborative learning among students.

3. Content-based programs have a larger impact on student achievement than pedagogical programs, but the final impact of the program depends on the combination of the subject and the specific training focus.

In the literature review of content-based programs aimed at enhancing teachers' knowledge of the subject, it was established that 67% of the estimates of the programs had positive and significant results; in contrast, of the pedagogical programs intended to improve teachers' pedagogical skills, 50% of the estimates had a positive and significant impact. These findings support those of Kennedy (1998), Garet et al. (2001), Clewell et al. (2004), Hill, Rowan, and Loewenberg (2005), and Boyd et al. (2009), suggesting that it is better to implement content-based programs when government intends to improve teachers' capabilities.

But this literature review has important implications in deciding public policies on the specific subjects of Math and Reading Comprehension. In particular, content-based programs are more likely to have a positive impact on Math performance. In Reading Comprehension, the use of pedagogical programs has a more positive impact. Justification for these mixed results can be found in Kennedy (1998), who argues that because of its characteristics (standardized testing, less teacher subjectivity, coordination across grades), Math is more suited to content-based training. In Science (which is similar to Reading Comprehension), teachers exercise a higher level of discretion in their teaching, and the curriculum is more flexible, which is more in line with the in-service programs that provides teachers with a combination of pedagogical behaviors and teaching knowledge.

The results of the impact evaluation in Chapter 3 are closer to the expected results. In terms of effect size, Reading Comprehension outperforms Math results in multigrade schools (0.148 vs 0.083 standard deviation) and full-grade schools (0.233 vs 0.137 standard deviation).

To explain these results for the Peruvian experience, a closer analysis is needed of the in-service teachers' program. According to the coaching protocols of the program (Ministry of Education, 2014), the coach works with teachers to improve and strengthen their pedagogical practices and school management based on critical and collaborative reflection. This interaction is carried out by the coach:

- promoting the effective use of time in the classroom;
- guiding teachers in optimizing the use of materials and resources available in the classroom (workbooks, texts, and classroom library);
- identifying and strengthening pedagogical strategies according to the sociocultural and linguistic context of the students;
- assessing teachers in the evaluation and analysis of student progress.

Also, the aforementioned protocols do not establish differences in the methodology used when training teachers in Mathematics or Reading Comprehension. Therefore, the training is more focused on improving teachers' pedagogical skills than on content-based training, which gives meaning to the results obtained.

4. There is a positive but decreasing effect of the duration of training on student outcomes: the findings suggest that the optimal length of training is 2 years.

As described in statement 2, there is a body of evidence on the positive impact of the duration of training programs in the educational sector. However, the literature review in Chapter 2 provides neither a clear framework for establishing the optimal duration of professional development programs nor confirmation as to whether any diminishing or increasing returns are caused by the long-term exposure to the intervention. What has been established so far are boundaries similar to those presented by Yoon et al. (2008), who states that programs with over 30 hours of intervention have positive results and Supovitz and Turner (2000), who illustrate the positive impact of 80 hours of intensive

training under the framework of the Local Systemic Change (LSC) initiative on teachers' behavior. The question about the optimal duration is even more essential in this thesis because the Peruvian program entails 169–224 training hours per year, more than similar programs evaluated by Yoon et al. (2008) and other Latin American experiences such as the Ecuadorian "Si Profe" or the Peruvian "Aprendes" program.

The results in Chapter 4 showed that, for the Peruvian program, the second year of intervention had diminishing but significant and positive returns in both Reading Comprehension (0.044 and 0.058 standard deviation) and Math (0.038 and 0.011 standard deviation) in multigrade and full-grade schools, respectively. Additionally, a third year of training had no positive nor significant impact on student achievement. As third-year results may be biased due to the small sample of schools (7 full-grade and 688 multigrade), a robustness analysis was conducted. The analysis confirmed those results. Because the second year of intervention – though decreasing – still had a positive and significant impact on student performance, the findings suggested that the optimal length of the program was 2 years.

The cause of these decreasing returns can be found in economic literature on a comparable sector used in this thesis, such as the labor sector. In this sector, an increase in input (labor) generally has a positive but decreasing marginal impact on final production. However, as Amoroso (2015) found, the exception to the rule occurred when the increase in labor input was accompanied by a specialization and knowledge component (human capital), which generates an increase in the marginal product.

Therefore, when analyzing Ministry of Education manuals for the Peruvian in-service teachers' program, it was observed that the training for each teacher, when exceeding one year, became repetitive. This repetition was because, for each additional year, the trained teacher received the same training content and same follow-up strategy as peers with one year of training. In the second year, the program focused on strengthening knowledge previously acquired, so there was no innovative component present in the second year. This explain the

theory of diminishing marginal returns in the Peruvian teacher training programs.

5. In the educational sector, on-the job training offers more effective results than off-the-job training.

The education sector has long debated how training should be delivered. Is it more effective for teacher training to be delivered within teachers' daily routine, including follow-up support through study groups, classroom demonstrations, and coaching (on-the-job training)? Or can traditional off-the-job training methodology be used; wherein general courses are given at a site outside of the working environment? Darling-Hammond and McLaughlin (2011) and Lieberman (1995) recommend that training must involve the teaching of concrete tasks, be based on the discussion of teachers' experiences (feedback) with students, and be given within the school environment to yield positive results. Also, Angrist and Lavy (2001), using a matching design, found that a teacher training program provided within schools in Jerusalem on a weekly basis had a positive impact on achievement.

In Chapter 5, a positive relationship was found between student achievement and the components of classroom visits and micro workshops, the two components closest to the definition of on-the-job training. On the other hand, the component related to off-the-job training methodology, update workshops, had a small or sometimes null effect.

As observed in Table 22, while the effects of the first two components accounted individually for approximately 40% of the impact of the program in Reading Comprehension and 10% in Math, update workshops had an almost null effect on results, with exception of Reading Comprehension in full-grade schools. To complement the previous findings, in Chapter 5-Table 20 showed that, when combining components such as classroom visits and micro workshops, results were greater than the sum of the individual effects for each component individually. This finding suggests a multiplicative effect if more than one component of the training is combined with others.

Table 22: The importance of each component in the final impact of the program in multigrade and full-grade schools.

	Multigrade schools %	Full-grade schools %
Classroom visits		
Reading	43%	43%
Math	14%	8%
Micro workshops		
Reading	39%	38%
Math	10%	5%
Update workshops		
Reading	2%	33%
Math	7%	0%

To better understand those differences, we examined the Peruvian program components in terms of the characteristics of both on- and off-the-job training studies and analyzed the relationship between the coach and the trained teacher. In this thesis we have described that classroom visits and workshops are close to the definition of on-the-job training because in these sessions, the coaches observe, give feedback, and discuss the classroom experiences with teachers, both individually and in groups, within the framework of what happens inside the classroom. Contrary, the update workshops, delivered not by the coach who knows the personal experience of each treated teacher, but instead by a general trainer that covers the entire district, include a range of general courses such as time management in the classroom, school management, pedagogical leadership, and school climate (Ministry of Education, 2014). These workshops are barely linked with the teachers'

environment and experiences, and so are closer to off-the-job training. Results confirm what has been found in the existing literature: on-the-job training components in the Peruvian in-service teachers' program, classroom visits, and micro workshops, provides better results than off-the-job training component, namely update workshops, on student achievement.

6. The Peruvian in-service teachers' program is cost effective compared to other international and national interventions also aimed at improving student results.

Chapter 3 found that the Peruvian in-service teachers' program had a positive and significant impact on student performance in Math and Reading Comprehension in both multigrade and full-grade schools, which is remarkable considering the particularities of the intervention (which is focused on rural areas with inequities). However, given this impact, it is important to determine if the intervention is cost effective, specifically whether the investment made was smaller than those made in interventions with similar results. To determine cost effectiveness, Yeh (2011) analyzes 22 different educational interventions that attempt to increase student achievement in the United States. Table 23 illustrates the cost per student in US$ in eight representative educational interventions and compares these costs with those of the Peruvian in-service teachers' program to determine cost effectiveness. To estimate the cost-effective index, the programs' costs were updated using the United States consumer annual price index (2006–2015).

Table 23: Cost-effective comparative analysis: Peruvian in-service training with other international interventions

Program/Intervention	Reading Comprehension (Cost per student in US$ for 0.1 SD)	Math (US$) (Cost per student in US$ for 0.1 SD)
Rapid Assessment[35]	6	7
Comprehensive school reform[36]	50	50
Longer school day	265	619
Teacher salary	510	510
Voucher program	35,027	13,838
Class size reduction[37]	1,541	1,781
10% increase in per pupil expenditure	1,566	1,566
High-quality preschool[38]	9,410	9,106
Peruvian In-service training for multigrade schools	**114.0**	**174.4**

Source: based on Yeh (2011)

The cost-effective results of Table 23 show that Peruvian in-service training (both multigrade and full-grade interventions), together with the Rapid Assessment, the Comprehensive school reform, and the Longer school day, were the more cost-effective interventions. These findings confirm what was previously proposed, that the Peruvian intervention could continue and, if possible, it should be scaled up to impact more teachers and students in all

[35] Program designed to allow students to read books according to their own level of understanding and individualized tutoring from teachers
[36] Refers to the Expeditionary Learning Outward Bound, a program that emphasizes methodology of learning by doing, group work, in-depth studies of specific topics, and public presentations.
[37] Project STAR. Nye, Hedges, and Spyros. (2001)
[38] Perry Pre-school intervention.

targeted areas of the intervention. In spite of these results, it is important to consider social and economic differences between US and Peruvian programs comparing their cost effectiveness.

Additionally, this thesis also compared the Peruvian in-service teachers' program, when focused on multigrade schools, with another Peruvian intervention named "Jornada Escolar Completa - JEC" (Longer school day), whose objective was to improve student results by increasing 10 lecture hours weekly in all secondary schools in Peru. For this analysis, the impact conducted by the Aguero (2016) was used, which found that JEC had a positive and significant impact on student achievement in Math and Reading Comprehension of 0.233 and 0.185 standard deviation, respectively. This analysis also presented cost information about the intervention.

Table 24: Cost-effective comparative analysis: Comparing two Peruvian educational programs

Program/Intervention	Reading Comprehension (Cost per student in US$ for 0.1 SD)	Math (US$) (Cost per student in US$ for 0.1 SD)
Peruvian In-service training for multigrade schools	114.0	174.4
Longer school day (JEC)[39]	490	389

As observed in Table 24, the teacher in-service training program in multigrade schools is more cost effective than the JEC program. This difference in results is expected to be more pronounced in full-grade schools, possibly because the cost of the implementation of the in-service program in full-grade schools is

[39] Based on Aguero (2016)

always lower (per capita) than in multigrade because of the numbers of students attending the two types of schools.

Considering these results, it can be affirmed that the Peruvian in-service teachers' program is one of the most cost-effective programs compared to similar interventions, both within Peru and internationally. These findings support the implementation of this type of interventions on a national -focused-scale and in other Latin American and developing countries, as a way of improving the educational performance of vulnerable students.

6.3 Policy Recommendations

The Peruvian in-service teachers' program has been found to have a positive impact on second graders' results in Reading Comprehension and Math in multigrade and full-grade schools. These results enable it to achieve one of the principal objectives of the PELA Program, namely the improvement of student performance in primary school. Also, the results show that the second year of intervention, in addition to achieving positive results in students' achievement, allows for a reduction in the urban/rural achievement gap, which is relevant as the program focuses mainly on rural areas. Therefore, the first policy recommendation offered by this thesis is to continue investing in the in-service program as it generates a positive impact for students who live in vulnerable conditions. This recommendation will also allow the government to reduce future income inequalities within the country, one of its principal priorities.

The literature on professional development programs previously assessed has found that sustained training is a key factor in generating changes in teaching behavior and student performance. One of the characteristics of the Peruvian in-service teachers' program is that it is a sustained intervention with 169- 224 training hours per year, which according to the literature review generates a greater probability of obtaining positive results in more than 12 percentage points. The analysis of the Peruvian program found that 2 years of intervention (between 338 and 448 hours) is the optimal length because, as well as the first

year, the second-year of intervention also has positive, though diminishing, results. According to the results of this thesis, it is advisable not to continue with the third year of implementation as suggested by the Ministry of Education because that additional year does not have a positive impact on student performance. Therefore, the second policy recommendation is to shorten the training period to two years.

This thesis has determined that only two of the components of the program have had a positive impact, namely classroom visits and micro workshops. Both components are close to the definition of "on-the-job training" because they encourage a direct relationship between the coach and the teachers trained within the working environment and have a positive impact on student performance. On the other hand, the update workshops had almost no effect on student performance, which may be because these workshops prioritize training outside the working environment, contrary to what is recommended by Lieberman (1995) and Darling-Hammond and McLaughlin (2011). Therefore, the third policy recommendation is to eliminate the update workshops and reassign resources to strengthening the first two components, which privilege discussion and feedback among colleagues of their teaching experiences with students.

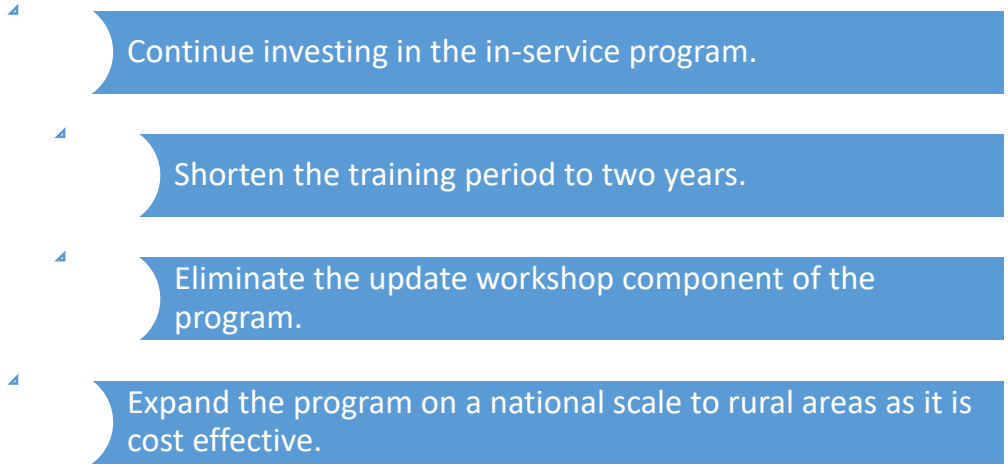
The in-service program is cost effective when compared to other Peruvian educational interventions such as the JEC program (increasing lecture hours to secondary students) and international programs aimed at improving student results. The cost of the Peruvian program ranges from 114 to 174 US$ for 0.1 standard deviation in multigrade schools, JEC is over 100% more expensive to get similar results. Internationally, compared to programs designed to provide individualized tutoring to students, which emphasize the methodology of learning by doing[40] or increasing class hours, the results are similar and sometimes more cost effective. Therefore, the fourth policy recommendation offered by this thesis is to expand the program on a national scale as it is a cost-

[40] Teaching method that consists of finding creative formulas and practices to generate learning based on real-life problems. This method implies that the teacher gives tools to students to encourage teamwork in practical activities.

effective tool that has enhanced student outcomes. Ideally, the government should aim to cover the 63% of schools in the target population that have not yet received intervention.

The four-policy recommendations for improving the design of the training program are summarized in Figure 16.

Figure 16: Policy recommendations to improve the Peruvian in-service teachers' program



6.4 Study limitations

This study has a number of limitations. The first of these limitations refers to the replicability of results at a national level. The results of this thesis are applicable to areas of high income inequalities, particularly rural areas with multigrade schools. Even if information and statistically significant results are available for full-grade schools, these data are usually found in urban low-income areas due to the focus of the program. This tendency may be observed in the characteristics of the treated and control samples, where there are significant

differences in variables related to poverty and rurality level. For this reason, this thesis used the PSM technique to develop the impact evaluation. Therefore, if the government intends to expand the program at a national level to include urban full-grade schools, more research is necessary to control for other variables that can affect (positively or negatively) the results in urban areas.

The second limitation of this thesis is that there are variables that could not be accounted for in the PSM models used but may explain part of the impact. In this thesis, the treated and non-treated sample were controlled by factors related to area income level, such as rurality, extreme poverty, inclusion in social programs, and malnutrition. These variables allowed us to meet the two conditions for obtaining unbiased results, namely conditional independence and common support. However, as mentioned by Supovitz and Turner (2000), it should be taken into account that teachers from poorer schools tend to use more traditional teaching practices than their urban counterparts due to the particularities of teaching in vulnerable environments, such as class size, available time, etc. Although the results obtained were statistically significant, future research should use variables that approximate these differences in teaching styles between urban and rural areas in order to obtain a more accurate model that reflects the final impact of the program.

The third limitation of this thesis refers to the number of results per study for the 21 papers selected for the literature review. Regardless of size, sample, degree of urbanity, and type of professional development, all the results – even if more than one result was produced per study – have been considered equally and accounted for in determining the final impact. However, due to the use of different methodologies, more than one result can be obtained per study. When considering all results, one study could be over represented, carrying more weight than the other studies in the final average impact. In the present thesis, studies by Angrist and Lavy (2001), Jacob and Lefgren (2004), Bressoux, Kramarz, and Prost (2008), Telese (2012), Monazza (2003), and Glazerman et al. (2010), with four results each, could be susceptible to overrepresentation.

6.5 Suggestions for future research

For future research, this thesis suggests further investigation into the impact of professional development programs in developing economies. From the 21 selected papers, this study was able to find only 4 studies delivered on developing countries that fulfilled the evidence standard requirements of this thesis. This thesis considered papers that fulfill an inclusion criterion elaborated over the basis of selected papers/guidelines (i.e., Van Klaveren and De Wolf (2014) and the Campbell Collaboration) that take into consideration specific requirements. These requirements are as follows: considering student achievement as the dependent variable; using Randomized Control Trial and Observational studies to assess the equivalence between the treated and control groups; including published papers; and collecting reliable data. Because of the scarcity of available papers, it was not possible to make inferences about how this type of program would function in other developing economies or to estimate whether such a program would attain different results than in developed countries.

Another topic that merits further research is the impact of the professional development program on secondary students. The literature review found that the impact of this type of program was ambiguous, with a similar probability (30%) of having significant positive and negative impacts. This topic should be carefully analyzed because it would be valuable to determine what caused the disparity in results between primary and secondary institutions. This disparity may be because the methodology used in training is similar to that used in primary education, though these institutions require more specialized training. Determining the cause for the difference in results may help developing countries to continue strengthening secondary education in order to ensure that vulnerable students can receive education of a higher standard and have a greater probability of accessing higher education.

Additionally, future research should examine the relationship between the quality of the coach and knowledge improvements in teachers. How much do

differences in coach quality affect teachers? In the initial periods of the Peruvian program (2008–2012), the selection of trainer (those in charge of training the coaches) was the responsibility of regional governments, which caused problems in terms of homogeneity in the quality of the training (Rodriguez, Sanz, and Soltau, 2013). To address this issue, the Ministry of Education decided in 2012 that all trainers are to be hired by the Ministry, while the hiring of coaches is to be left to regional governments. This decision, which homogenized the quality of trainers and coach training, may be the basis of a future study that seeks to determine the differentiated impact of the quality of coach quality on teachers' knowledge. If academia continues exploring this relationship, research will be able to understand more fully the virtuous circle of coach-teacher-student, which may aid in understanding teacher training programs' impact on students and students' educational outcomes.

About the persistence of the results, the fourth chapter evaluated how the annual results of the students increase when the school has been intervened for more than one year. However, we know little about the persistence of positive impacts over time. Could it be that the two-year training programs, even if they have decreasing annual impacts, can improve the duration of the positive effects of the first year on students in the medium and long term? Or even that the schools treated for only one year remain with the positive effects for the following cohort of students? Therefore, a comparative analysis over time between schools that have received different intensities of treatment would help us to see the persistence of the program over time.

Finally, this thesis recommends testing the explanations provided by this thesis to explain the reasons for the varying impact of the Peruvian in-service teachers' program. Are the differences in multigrade schools and the effect of peers the reasons for the lesser impact of the Peruvian program compared to the impacts of similar programs elsewhere in the world? Is it possible to improve Math performance if policy-makers apply a differentiated content-based strategy in teacher trainings in Math? If different content was generated for the second or third years of the training, could the decreasing returns be reversed, and could there even be a positive impact in the third year of intervention? The answers

to these questions would not only serve the Peruvian government in improving the implementation of its program, but also the international community in improving the results of professional development programs.

Bibliography

1. Aguero, J. (2016). Impact evaluation of Jornada Escolar Completa-JEC. *FORGE.* Retrieved from http://www.grade.org.pe/forge/descargas/Evaluación%20de%20impacto%20de%20la%20jornada%20escolar%20completa.pdf

2. Amoroso, S. (2015). Profits, R&D and labour. Breaking the law of diminishing returns to labour. (IPTS Working Papers on Corporate R&D and Innovation – No 10/2015). Retrieved from https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/profits-rd-and-labour-breaking-law-diminishing-returns-labour

3. Anderson, J. B. (2000). Factors affecting learning of Mexican primary school children. *Estudios Economicos, 15 (1)*, 117-152. Retrieved from https://www.jstor.org/stable/40311428?seq=1#page_scan_tab_contents

4. Angrist, J., & Lavy, V. (2001). Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools. *Journal of Labor Economics, 19 (2)*, 343-369. Retrieved from https://economics.mit.edu/files/11103

5. Alcazar, L & Guerrero, L. (2011). Improving Basic Education in Peru: Review of major educational programs. Retrieved from http://www.grade.org.pe/publicaciones/1057-mejora-efectiva-de-la-educacion-basica-en-las-zonas-rurales-del-peru/

6. Araujo, M., Carneiro, P., Cruz-Aguayo, Y. & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics, 131 (3)*, 1415-1453.

7. Bacolod, M. & Tobias, J. (December 2006). Schools, school quality and achievement growth: Evidence from the Philippines. *Economics of Education*

Review, 25 (6), 619-632. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0272775705000919

8. Bando, R. & Li,X. (2014). The Effect of In-Service Teacher Training on Student Learning of English as a Second Language, *IDB Working Paper Series*, *529*. Retrieved from https://publications.iadb.org/handle/11319/6596

9. Bedard, K. & Ferrall C. (2003). Wage and Test Score Dispersion: Some International Evidence. *Economics of Education Review, 22(1)*, 31–43. https://doi.org/10.1016/S0272-7757(01)00060-7

10. Benavides, M., Diaz, J. & Villaran, V. (2009). *Evaluation of the distribution of educational materials of the Ministry of Education*. Retrieved from http://repositorio.minedu.gob.pe/bitstream/handle/123456789/893/521.%20Presupuesto%20evaluado%20Materiales%20educativos..pdf?sequence=1&isAllowed=y

11. Bernhaum, M., Rivero, J & Schiefelbein, E. (2010). *Evaluation of USAID/Peru's education program: APRENDES AND CETT-Andino. Final Evaluation Report*. Retrieved from https://es.scribd.com/document/374774788/Evaluacion-de-Los-Programas-Educativos-de-USAID-PERU-AprenDes

12. Biewen, M., Fitzenbergen, B., Osikominu, A., Waller, M. (2007). *Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany. Institute for the Study of Labor (IZA) Discussion Papers 2885*. Retrieved from http://ftp.iza.org/dp2885.pdf

13. Blau, F. & Kahn, L. (2005). Do Cognitive Test Scores Explain Higher US Wage Inequality? *Review of Economics and Statistics, 87 (1)*, 184–193. Retrieved from https://www.jstor.org/stable/40042932?seq=1#page_scan_tab_contents

14. Blundell, R., & Costa Dias, M. (2009). Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources, 44 (3),* 565-640. https://doi: 10.3368/jhr.44.3.565

15. Bossiere, M. (2004). *Determinants of primary education outcomes in developing countries. World Bank Operations Evaluation Department (OED) paper series, 39157*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.2919&rep=rep1&type=pdf

16. Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31(4*), 416-440. https://doi.org/10.3102/0162373709353129

17. Bruns, B., Costa, L., & Cunha, N. (2018). Through the Looking Glass: Can Classroom Observation and Coaching Improve Teacher Performance in Brazil? *Economics of Education Review, 64, June*, 214-250. https://doi.org/10.1016/j.econedurev.2018.03.003

18. Card, D., Kluve, J., & Weber, A. (2010). Active Labour Market Policy Evaluations: A Meta-Analysis. *The Economic Journal, 120,* F452-F477. Retrieved from http://davidcard.berkeley.edu/papers/card-kluve-weber-EJ.pdf

19. Carr-Hill, R., Rolleston, C., Pherali, T., & Schendel, R. (2015). The effects of school- based decision making on educational outcomes in low and middle-income contexts: a Systematic Review. *International Initiative for Impact Evaluation (3ie), September 2015*. Retrieved from http://www.3ieimpact.org/media/filer_public/2015/11/20/dfid-funded-decentralisation-review.pdf

20. Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499-53. Retrieved from https://www.jstor.org/stable/1162862

21. Carrillo, C., Maassen van den Brink, H., & Groot, W. (2016). *Professional development programs and their effects on student achievement: A systematic review of the evidence*. *TIER WORKING PAPER SERIES, 16/03*. Retrieved from http://www.tierweb.nl/home/assets/files/UM/Working%20papers/TIER%20WP%2016-03.pdf

22. Carrillo, C., Maassen van den Brink, H., & Groot, W. (2017). *The Peruvian teacher in-service training program and its effects on student achievement. TIER WORKING PAPER SERIES, 17/14*. Retrieved from http://www.tierweb.nl/tier/assets/files/working_papers/TIER%20WP%2017-14.pdf

23. Carrillo, C., Maassen van den Brink, H., & Groot, W. (2017). *The sustainability of the impact of the Peruvian teacher in-service training program on student achievement over time. TIER WORKING PAPER SERIES*.

24. Carrillo, C., Maassen van den Brink, H., & Groot, W. (2018). *The effectiveness of different components of a teacher in-service training program in Peru. TIER WORKING PAPER SERIES*.

25. Caulfield-Sloan, M & Ruzicka, M. (2005). The effect of staff development of teachers in the use of higher order questioning strategies on third grade students' rubric science assessment performance. *Planning and Changing, 36 (3-4),* 157-175. Retrieved from https://files.eric.ed.gov/fulltext/EJ737676.pdf

26. Checchi, D., & van de Werfhorst, H. G. (2017, February 9). Policies, Skills and Earnings: How Educational Inequality Affects Earnings Inequality. https://doi.org/10.31235/osf.io/wn3h2

27. Chingos, M.M., & Peterson, P.E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30 (3),* 449 – 465. https://doi.org/10.1016/j.econedurev.2010.12.010

28. Ciccone, A., & Papaioannou, E. (2009). Human capital, the structure of production, and growth. *Review of Economics and Statistics, 91(1),* 66–82. https://doi.org/10.1162/rest.91.1.66

29. Cohen, D.K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record,*

102 (2). Retrieved from http://www-personal.umich.edu/~dkcohen/cohen_hill_2000_TCR.pdf

30. Clewell, B.C., Campbell, P.B., & Perlman, L. (2004). Review of Evaluation Studies of Mathematics and Science Curricula and Professional Development Models. *GE Foundation*. Retrieved from https://www.urban.org/sites/default/files/publication/51796/411149-Review-of-Evaluation-Studies-of-Mathematics-and-Science-Curricula-and-Professional-Development-Models.PDF

31. Cohen, D.K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102 (2), 294–343*. Retrieved from http://www-personal.umich.edu/~dkcohen/cohen_hill_2000_TCR.pdf

32. Conti G. (2005). Training, productivity and wages in Italy. *Labour Economics, 12 (4),* 557–576. https://doi.org/10.1016/j.labeco.2005.05.007

33. Crépon, B., Ferracci, M., & Fougère, D. (2012). Training the Unemployed in France: How Does it Affect Unemployment Duration and Recurrence? *Annals of Economics and Statistics, 107/108*, 175-199. Retrieved from https://www.jstor.org/stable/23646576

34. Darling-Hammond, L., & McLaughlin, M.W. (2011). Policies that support professional development in an era of reform. *Phi Delta Kappa, 92 (6),* 81-92. https://doi.org/10.1177/003172171109200622

35. Darling-Hammond, L., Wei, R., Andree, A., Richardson, N., Orphanos, S. (2009). National Staff Development Council and The School Redesign Network at Stanford University Working paper. Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad. National Staff Development Council. Retrieved from https://learningforward.org/docs/default-source/pdf/nsdcstudy2009.pdf

36. Dearden L., Reed H., Van Reenen J. (2006). The impact of training on productivity and wages: Evidence from British panel data. *Oxford Bulletin of*

Economics and Statistics, 68, 397–421. https://doi.org/10.1111/j.1468-0084.2006.00170.x

37. De Grip, A., and J. Sauermann (2012), The Effects of Training on Own and Co-Worker Productivity: Evidence from a Field Experiment. *Economic Journal, 122(560)*, 376-399.

38. Del Mastro, C. (2008). *Assessment of the budget execution of the National Program for Continuing Education and Training (PRONAFCAP)*. National Budget Office (DNPP). Retrieved from https://www.mef.gob.pe/contenidos/presu_publ/ppr/eval_indep/2008_informe_final_PRONAFCAP.pdf

39. Díaz, J. J, & Ñopo, H. (2017). *The teaching career in Peru. GRADE*. Retrieved from http://www.grade.org.pe/publicaciones/la-carrera-docente-en-el-peru/

40. Education quality agency-Chile. (2018). *National students results 2017*. Retrieved from http://archivos.agenciaeducacion.cl/PPT_Conferencia_ER_2017_web_3.pdf

41. Educational reform: progress and pending challenges. *Economic Friday, PUCP. 29 april 2016*. Retrieved from https://slideplayer.es/slide/10492695/

42. Epple, Dennis, & Richard Romano. (2011). Peer Effects in Education: A Survey of the Theory and Evidence. *Handbook of Social Economics, 1B*, 1053–1163. https://doi.org/10.1016/B978-0-444-53707-2.00003-7

43. Fitzenberg, B., Osikominu, A., & Paul, M. (2010). The heterogeneous effects of training incidence and duration on labor market transitions. *Centre for European Economic Research (ZEN) Paper 10-077*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1726424

44. Fitzenberger, B., Osikominu, A., & Robert, V. (2006). Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed

in West Germany. *ZEW Discussion Paper No. 06-039*. Retrieved from ftp://ftp.zew.de/pub/zew-docs/dp/dp06039.pdf

45. Fitzenberger, B., Volter, R. February (2007). Long-Run Effects of Training Programs for the Unemployed in East Germany. *Labour Economics, 14 (4)*. https://doi.org/10.1016/j.labeco.2007.05.002

46. Garet, M., Wayne A., Stancavage, F., Taylor J., Walters K., Song M., Brown S., Hurbulrt S., Zhu P., Sepaink S., Doolittle F., & Warner E. (2011). Middle school mathematics professional development impact study: Findings after the second year of implementation. *National Center for Education Evaluation and Regional Assistance, 2011-4024, U.S. Department of Education*. Retrieved from https://files.eric.ed.gov/fulltext/ED519922.pdf

47. Garet, M., Cronen S., Eaton, M., Kurki A., Ludwig M., Wehman J., Kazuaki U., Audrey F., Bloom H., Zhu P., Sztejnberg L., & Doolittle F. (2008). The impact of two Professional Development interventions on early reading instruction. *National Center for Education Evaluation and Regional Assistance, 2008-4030, U.S. Department of Education*. Retrieved from https://ies.ed.gov/ncee/pdf/20084030.pdf

48. Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38(4)*, 915–945. https://doi.org/10.3102/00028312038004915

49. Geoffrey B. Saxe, Maryl Gearhart, Na'ilah Suad Nasir. (2001). Enhancing Students' Understanding of Mathematics: A Study of Three Contrasting Approaches to Professional Support. *Journal of Mathematics Teacher Education. Volume 4 (1)*, pp 55-79. Retrieved from https://link.springer.com/article/10.1023/A:1009935100676

50. Glatthorn, A. (1995). Teacher Development. In L. Anderson (Ed), *International encyclopedia of teacher and teaching education. p.41*. London, Pergamon Press.

51. Glazerman S., Isenberg E., Dolfin S., Bleeker M., Johnson A., Grider M., Jacobus M., & Ali M. (2010). Impacts of Comprehensive Teacher. *National Center for Education Evaluation and Regional Assistance, 2010-4027, U.S. Department of Education*. Retrieved from https://files.eric.ed.gov/fulltext/ED565837.pdf

52. Glewwe, Paul, Kremer, Michael, Moulin Sylvie, & Zitzewitz, Eric. (2004). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics 74 (1) (Special Issue June 2004),* 251-68. https://doi.org/10.1016/j.jdeveco.2003.12.010

53. Goldschmidt, P., & G. Phelps. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review, 29,* 432 – 439. https://doi.org/10.1016/j.econedurev.2009.10.002

54. Guadalupe, C., León, J., Rodríguez, J.S., Vargas, S. (2017). Situation of education in Peru. basic education analysis and perspectives. *GRADE*. Retrieved from http://www.grade.org.pe/forge/descargas/Estado%20de%20la%20educación%20en%20el%20Perú.pdf

55. Haelermans, C., & Borghans, L. (2011). Wage effects of on-the-job training: A Meta-Analysis. *Institute for the Study of Labor (IZA) Discussion Paper 6077*. Retrieved from http://ftp.iza.org/dp6077.pdf

56. Hanushek, Eric. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data. *American Economic Review 61(2),* 280-288. Retrieved from http://hanushek.stanford.edu/publications/teacher-characteristics-and-gains-student-achievement-estimation-using-micro-data

57. Hanushek, Eric. (1995). Interpreting recent research on schooling in developing countries. *World Bank Research Observer, 10 (2),* 227-243. Retrieved from https://www.jstor.org/stable/3986584?seq=1#page_scan_tab_contents

58. Hanushek, Eric, & Ludger Woessmann. (2007). The Role of Education Quality in Economic Growth. *World Bank Policy Research Working Paper 4122*. https://doi.org/10.1596/1813-9450-4122

59. Hasan Ünal, I. D. (2011). Teachers' Professional Development and Students' Mathematics Performance: Findings from TIMSS. Procedia - Social and Behavioral Sciences, 15, 3252-3257. https://doi.org/10.1016/j.sbspro.2011.04.280

60. Harbison, R.W., & Eric Hanushek, E.A. (1992). Educational performance for the poor: Lessons from rural northeast Brazil. *Oxford University Press*. Retrieved from http://documents.worldbank.org/curated/en/310981468770471105/Educational-performance-of-the-poor-lessons-from-rural-northeast-Brazil

61. Harris, D. & Sass, T. (2011). Teacher Training, Teacher Quality, and Student Achievement. Journal of Public Economics, 95 (7–8), 798-812. https://doi.org/10.1016/j.jpubeco.2010.11.009

62. Harris, D., & Sass, T. (2009). What makes for a good teacher and who can tell? *National Center for Analysis of Longitudinal Data in Education Research-CALDER, working paper 30*. Retrieved from https://www.urban.org/sites/default/files/publication/33276/1001431-What-Makes-for-a-Good-Teacher-and-Who-Can-Tell-.PDF

63. Harris, D., & Sass, T. (2001). The Effects of Teacher Training on Teacher Value Added. *Journal of Teacher Education, 51*, 164-169. Retrieved from https://www.researchgate.net/publication/5057696_The_Effects_of_Teacher_Training_on_Teacher_Value_Added

64. Hill, H., Rowan, B., & Loewenberg, D. (2005). Effects of Teachers Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal, 42 (2),* 371-406. Retrieved from http://lmt.soe.umich.edu/files/hillrowanball.pdf

65. Hunt, Barbara. (2009). Teacher Effectiveness: A Review of the International Literature and Its Relevance for Improving Education in Latin America. *Partnership for Educational Revitalization in the Americas (PREAL) Working Paper Series, 43*. Retrieved from http://archive.thedialogue.org/PublicationFiles/WP43%20-%20English%20(final).pdf

66. Hujer, R., Thomsen, L., & Zeiss, C. (2006). The Effect of Vocational Training Programmes on the Duration of Unemployment in East Germany. *Allgemeines Statislisches Archiv, 90* (2), 299-321. Retrieved from https://link.springer.com/article/10.1007/s10182-006-0235-z

67. Ibarraran, P., & D. Rosas. (December 2008). *Evaluating the Impact of Job Training Programs in Latin America: Evidence from IDB funded operations*. Retrieved from http://conference.iza.org/conference_files/ELMPDC2009/ibarraran_p4263.pdf

68. Jacob, B., & Lefgren, L. (2004). The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago. *Journal of Human Resources, 39 (1), 50-79*. Retrieved from https://www.jstor.org/stable/3559005

69. João Batista Gomes-Neto, Eric A. Hanushek. (1994). Causes and consequences of grade repetition: Evidence from Brazil. *Economic Development and Cultural Change*. *43(1)*, 117-148. Retrieved from http://hanushek.stanford.edu/publications/causes-and-consequences-grade-repetition-evidence-brazil

70. Joshua D. Angrist & Jörn-Steffen Pischke. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. *Princeton University Press*. Retrieved from https://www.researchgate.net/publication/51992844_Mostly_Harmless_Econometrics_An_Empiricist%27s_Companion

71. Kang, C. (2007). Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea. *Journal of Urban Economics, 61*, 458–495. https://doi.org/10.1016/j.jue.2006.07.006

72. Khattak Z,I., & Abbasi, M.G. (2010). Evaluation of the effectiveness of in-service teacher training courses of the CALL sub-committee of the ELTR project in Pakistan. *Procedia Social and Behavioral Sciences, 2,* 4911-4917. https://doi.org/10.1016/j.sbspro.2010.03.794

73. Kennedy, Mary. (1998). Form and Substance in In-service Teacher Education. *National Institute for Science Education, Research Monograph, 13.* Retrieved from http://archive.wceruw.org/nise/Publications/Research_Monographs/vol13.pdf

74. Kretlow, A., Cooke, N., & Wood, C. (2012). Using In-Service and Coaching to Increase Teachers' Accurate Use of Research-Based Strategies. SAGE, 13 (6). https://doi.org/10.1177/0741932510395397

75. Kluve, J. (2006). The Effectiveness of European Active Labour Market Policy. *Institute for the Study of Labor (IZA) Discussion Paper 2018*. Retrieved from http://ftp.iza.org/dp2018.pdf

76. Kluve, J., Schneider, A., Uhlendorff, A., & Zhao, Z. (2007). Evaluating Continuous Training Programs Using the Generalized Propensity Score. *Institute for the Study of Labor (IZA) Discussion Paper 3255*. https://doi.org/10.1111/j.1467-985X.2011.01000.x

77. King, M., & Behrman, J. (2009). Timing and duration of exposure in evaluation of social program. *Policy Research Working Paper No. 4686. World Bank*. Retrieved from https://openknowledge.worldbank.org/handle/10986/6812

78. Lazarová, Bohumíra & Alice Prokopová. (2004). *Psychological aspects of in-service teacher training: educational needs and expectations*. Paper presented at the European Conference on Educational Research, University of

Crete, 22-25 September 2004. Retrieved from http://www.leeds.ac.uk/educol/documents/00003630.htm

79. Lieberman, A. (1995). Practices that support teacher development. *Phi Delta Kappan 76, April 1995*. 591-596. Retrieved from https://eric.ed.gov/?id=EJ501258

80. Liebowitz, Arlecn. (1974). Home investments in children. *Journal of Political Economy, 82(2)*, 111-131. Retrieved from https://econpapers.repec.org/article/ucpjpolec/v_3a82_3ay_3a1974_3ai_3a2_3ap_3as111-s131.htm

81. Loucks-Horsley, S. & Matsumoto, C. (1999). Research on Professional Development for Teachers of Mathematics and Science: The State of the Scene. *School Science and Mathematics, 99 (*5), 258–271. https://doi.org/10.1111/j.1949-8594.1999.tb17484.x

82. Majerowicz, M. & R. Montero. (2018). *Working paper.* Can Teaching be Taught? Experimental Evidence from a Teacher Coaching Program in Peru. Retrieved from https://scholar.harvard.edu/files/coaching.pdf

83. Mason, G., B. O'Leary, & M. Vecchi. (2012). Certified and uncertified skills and productivity growth performance: Cross-country evidence at industry level. *Labour Economics, 19*. 351 – 360. Retrieved from https://www.researchgate.net/publication/254409365_Certified_and_Uncertified_Skills_and_Productivity_Growth_Performance_Cross-Country_Evidence_at_Industry_Level

84. McCollum, J., Hemmeter, M.L., Hsieh, W. (May 2013). Coaching Teachers for Emergent Literacy Instruction Using Performance-Based Feedback. *Topics in Early Childhood Special Education*, 33 (1), 28-37. Retrieved from https://eric.ed.gov/?id=EJ998151

85. McCutchen, D., Abbot, R.D., Green, L.B., Beretvas, S.N., Cox, S., Potter, N.S., Quiroga T., and Gray AL. (2002). Beginning. Literacy: Links among Teacher

Knowledge, Teacher Practice, and Student Learning. *Journal of Learning Disabilities*. 69-86. https://doi.org/10.1177/002221940203500106

86. McGill-Franzen, A., Allington, R. L., Yokoi, L., & Brooks, G. (1999). Putting books in the classroom seems necessary but not sufficient. *Journal of Reading Research, 93(2)*, 67–74. https://doi.org/10.1080/00220679909597631

87. McGuinness, S., O'Connel, P., & Kelly, E. (2014). The Impact of Training Programme Type and Duration on the Employment Chances of the Unemployed in Ireland. *Economic and Social Review, 45 (3)*, 425-450. Retrieved from https://www.esr.ie/article/view/189

88. Metzler, J., & L. Woessmann. (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. https://doi.org/10.1016/j.jdeveco.2012.06.002

89. Miranda, Liliana. (2008). Factors associated with school performance and their implications for educational policy in Peru. In Benavides, M. Analysis of programs, processes and educational outcomes in Peru. Empirical contributions to the debate. *GRADE*. Retrieved from http://www.grade.org.pe/upload/publicaciones/archivo/download/pubs/analisis0.pdf

90. Ministry of Economics and Finance. (2010). *PELA presentation 2010*. Retrieved from https://www.mef.gob.pe/contenidos/presu_publ/documentac/pat_educacion.pdf

91. Ministry of Economics and Finance. (2013). *Roundtable for poverty reduction. Monitoring the Budget Program of Learning Achievement (PELA): Balance for the year 2012*.

92. Ministry of Economics and Finance. (2015). *Unified Financial Management System (SIAF)*. Retrieved from http://apps5.mineco.gob.pe/transparencia/Navegador/default.aspx

93. Ministry of Economics and Finance. (2013). *Policy Brief 2013*. Retrieved from https://www.mef.gob.pe/es/presupuesto-publico-sp-18162/624-evaluacion-independiente/5387-sector-educacion

94. Ministry of Education. (2012). *Balance for the in-service Program within PELA. Department of Regular Education*. Retrieved from https://www.mef.gob.pe/contenidos/presu_publ/ppr/talleres/ppat2013/04julio2012/1EDUCACION/PELA.pdf

95. Ministry of Education. (2012). *Teacher in-service program Report from SIGMA*. *General Direction of Regular Education*. Retrieved from https://sigma2.minedu.gob.pe

96. Ministry of Education. (2014). *In-service Program Protocol*. Retrieved from http://repositorio.minedu.gob.pe/bitstream/handle/123456789/3706/Protocolo%20de%20acompañamiento%20pedagógico.pdf?sequence=1&isAllowed=y

97. Ministry of Education-PELA Annex 2. (2015). *The PELA Program 2016-2018: Annex 2 for the results-based budgeting program*.

98. Ministry of Education. (2016). *Educational Reform, Advances and pending challenges*. *Economic Friday, PUCP, 29 april 2016*. Retrieved from https://slideplayer.es/slide/10492695/

99. Ministry of Education. (2016). *2011-2016 compliance report*. Retrieved from http://www.minedu.gob.pe/transferencia-de-gestion/pdf/reporte-del-cumplimiento-misional1.pdf

100. Ministry of Education. (2016). *Infrastructure census results*. Retrieved from http://www.minedu.gob.pe/censo-infraestructura-educativa/

101. Ministry of Education. (2016). *ECE report*. Retrieved from http://umc.minedu.gob.pe/resultadosece2016/

102. Ministry of Education. 2017. *Peru in PISA 2015*. Retrieved from http://umc.minedu.gob.pe/wp-content/uploads/2017/04/Informe-PISA-2015_ALTA.pdf

103. Monazza, A. (2000). The determinants of student achievement in government and private schools in Pakistan. *Pakistan Development Review, 42 (4),* 841-75. Retrieved from https://www.jstor.org/stable/41260442

104. Navarro, J. C. & Verdisco, A. (2010). Teacher training: What works and what does not. Innovations and trends in Latin America. Retrieved from http://www.educarchile.cl/

105. Nye, B., Hedges, L., Konstantopoulos, S. (2001). Are Effects of Small Classes Cumulative? Evidence from a Tennessee Experiment. *The Journal of Educational Research, 94 (6)*, 336-345. https://doi.org/10.1080/00220670109598771

106. Bressoux, P., Kramarz, F., & Prost, C. (2008). Teachers' Training, Class Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes. *The Economic Journal, 119, (536),* 540-561. https://doi.org/10.1111/j.1468-0297.2008.02247.x

107. Podhajski, B., Mather, N., Nathan, J., & Sammons, J. (2009). Professional Development in Scientifically Based Reading Instruction: Teacher Knowledge and Reading Outcomes. *Journal of Learning Disabilities, 42 (5)*, 403-441. https://doi.org/10.1177/0022219409338737

108. Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica, 73(2)*, 417-458. Retrieved from https://econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf

109. Robbins, S., De Cenzo, D., & Coulter, M. (2016). *Fundamentals of Management: Essential Concepts and Application*s. Pearson (9 edition).

110. Rodriguez, J., Sanz, P., & Soltau, L. (2013). Evaluation of design and budget implementation: teacher in-service program. *Ministry of Education*. Retrieved from https://www.mef.gob.pe/contenidos/presu_publ/ppr/eval_indep/2013_pedagogico.pdf

111. Rowan, B., Chiang, F., & Miller, R.J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education, 70*, 256-284. Retrieved from https://www.jstor.org/stable/2673267

112. Rosangela Bando & Xia Li. (2014). The Effect of In-Service Teacher Training on Student Learning of English as a Second Language, *Inter-American Development Bank Working Papers 86173*. Retrieved from https://publications.iadb.org/handle/11319/6596

113. Rosenbaum, P., & Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika, 70 (1),* 41-55. Retrieved from http://www.stat.cmu.edu/~ryantibs/journalclub/rosenbaum_1983.pdf

114. Supovitz, Jonathan & Herbert Turner. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*, 963-980. Retrieved from https://pdfs.semanticscholar.org/f3fb/cc1b42f78fab755e85135ef2ab516fe56529.pdf

115. Swinton, J., Scafidi, B., & Woodard, H. (2012). The Impact of the Teaching High School Economics Workshop for Teachers on Student Achievement. *Eastern Economic Journal, 38*, 401-416. Retrieved from https://jswinton.gcsu.edu/Research/Interests_files/teaching%20HS%20Economics.pdf

116. Telese, J. A. (2012). Middle School Mathematics Teachers' Professional Development and Student Achievement. *Journal of Educational Research, 2,* 102-111. https://doi.org/10.1080/00220671.2010.521209

117. Tirivayia, N., Maasen van den Brink, H. & Groot, W. (2013). Group incentives for teachers and their effects on student learning: a systematic review of theory and evidence. School Effectiveness and School Improvement*, 25 (4)*. https://doi.org/10.1080/09243453.2013.857697

118. UNESCO. (2010). Building the Wealth of Nations. *World Conference on Care and Early Childhood Education. Moscow*. Retrieved from http://www.unesco.org/education/WCECCE/ConceptPaper.pdf

119. Van Klaveren, C. & De Wolf, I. (2013). Systematic Reviews in Education Research: When Do Effect Studies Provide Evidence?, *Top Institute for Evidence Based Education Research Working Papers Series, TIER WP 13/02*. Retrieved from http://www.tierweb.nl/assets/files/UM/Working%20papers/When%20do%20Effect%20Studies%20Provide%20Evidence%20Working%20Paper.pdf

120. Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to intervention as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69*, 391–409. https://doi.org/10.1177/001440290306900401

121. Vooren, M., Haelermans, C., Groot, W., Maassen van den Brink, H. (2018). The effectiveness of active labour market policies: a meta-analysis. *Journal of Economic Surveys,* 0 (0), 1–25. https://doi.org/10.1111/joes.12269

122. Woessmann, Ludger. (2001). Schooling Resources, Educational Institutions, and Student Performance: The International Evidence. *Oxford Bulletin of Economics and Statistics, 65 (2),* 117-170. https://doi.org/10.1111/1468-0084.00045

123. World Bank. *Peru: Systematic Country Diagnostic 2017*. Retrieved from http://documents.worldbank.org/curated/en/919181490109288624/pdf/Peru-SCD-final-3-16-17-03162017.pdf

124. World Bank-DataBank. (2018). *Government expenditure on education, total (% of GDP)*. Retrieved from https://datos.bancomundial.org/indicador/SE.XPD.TOTL.GD.ZS?end=2015&locations=PE-ZJ&start=2001&view=chart

125. Yeh, Stuart S. (2010). The Cost Effectiveness of 22 Approaches for Raising Student Achievement. *Journal of Education Finance, 36 (1)*, 38-75. Retrieved from https://www.jstor.org/stable/40704405

126. Yoon, K. S., Garet, M., Birman, B. & Jacobson, R. (2007). Examining the effects of mathematics and science professional development on teachers' instructional practice: Using professional development activity log. *Council of Chief State School Officers, Washington DC*. Retrieved from http://programs.ccsso.org/content/pdfs/AIR%20MSP_report_Final%20021307.pdf

127. Yoon, K. S., Duncan, T., Lee, S., Scarloss, B. & Shapley, K. (2007). Yoon, K. S., Duncan, T., Lee, S. & Shapley, K. (2008). Reviewing the evidence on how teacher professional development affects student achievement. *Regional Educational laboratory at advance research REL 2007, 033*. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf

128. Yoon, K. S., Duncan, T., Lee, S. & Shapley, K. (2008). The effects of teachers' professional development on student achievement: Findings from a systematic review of evidence. *Paper presented at the Annual Meeting of the American Educational Research Association, New York*.

129. Zan, B. & Donegan-Ritter, M. (2014). Reflecting, Coaching and Mentoring to Enhance Teacher–Child Interactions in Head Start Classrooms. *Early Childhood Education Journal*, *42 (2)*, 93–104. Retrieved from https://eric.ed.gov/?id=EJ1036162

Summary

Professional development programs, which usually take the form of personal coaching, conferences and workshops, are a key element to improve the capacities of teachers. Teachers, in turn, are a primary factor in understanding how students achieve learning outcomes. Therefore, it is important to measure in a quantitative manner the impact that professional development programs have on student achievement.

For a developing economy like Peru, this analysis is especially important. Peru's average annual growth rate of 5.4% for the period 2001-2017 was not accompanied by a reduction of income inequalities or by substantial improvement of students' academic performance. The country's position on PISA test results (67th out of 75 countries/regions in Math, and 68th of 75 countries/regions in Reading Comprehension), together with a considerable gap between urban and rural students, has made the challenge of improving education a national priority. This priority is even more so the case if the country accepts Hanushek and Woessmann's (2012) argument that student achievement is correlated with the sustainability of long-term economic development.

The Peruvian government has committed to increasing student results and reducing the inequities between urban and rural areas. Accordingly, in 2008, the Peruvian government launched the PELA program, a strategy intended to improve the results of students aged 3-16 years old. The strategy aims to address areas related to school infrastructure, teacher development, educational material, and monitoring and evaluation. The strategy has increased so much that in 2016, it occupied 58% of the Peruvian educational budget.

The objective of this study is to quantitatively assess the impact on student results of a specific program included in the second PELA area (teacher development): the Peruvian teachers' in-service program. This intervention focuses mainly on improving the pedagogical practices of teachers working in the context of high-income inequality. The intervention operates through a systematic process of conducting meetings between a coach and teachers and

presenting several workshops. The importance of the in-service program is such that during 2016, it covered 8,726 schools throughout the country. In the future, it is planned to be expanded to 63% of the vulnerable schools that have not yet been covered.

The most recent study on the impact of in-service training programs on student achievement, that also used a clear methodology and rigorous standards to systematically compare results, was conducted 10 years ago (Yoon et al., 2008). Yoon et al.'s (2008) study is part of a series of systematic studies that started in the late 1990s with the one carried out by Kennedy et al. (1998) who based their results on experimental or quasi-experimental studies who works with reliable databases. These research's show that these training programs do have a positive impact on student performance.

However, most of the studies included in those systematic reviews have evaluated the relationship between in-service programs and student outcomes in the context of an urban environment in a developing economy. There is little information about the effects of these programs on students who live in vulnerable conditions like those in the Peruvian case.

Therefore, the added value of this thesis is its study of the impact of a teachers' in-service program that is focused on vulnerable students. For the purpose of this study, we have a unique dataset that includes the academic results of every second-grader of all public schools in Peru together with information on the socio-economic variables of the students and the level of compliance of each teacher with the components of the program during the period 2014-2016.

The main finding of the Peruvian experience is that the program has positive impacts on educational performance. This finding coincides with most of the findings presented in the systematic literature review of this thesis, which covers the last 20 years. Besides, the detailed analysis the program also provides new information about the characteristics that a teacher's in-service program should have in order to be more efficient among vulnerable students: Namely, the analysis explains how duration, intensity, type of coach-teacher contact, and characteristics of teaching have differentiated impacts on student performance.

Chapter 2 of this thesis studies the state-of-the art of the empirical research from the last 20 years on the impact of professional development programs on

student achievement. This chapter, using a search strategy and a specific inclusion criterion for the selection of studies, systematizes the impacts of public and private interventions over three regular educational levels: pre-school, elementary, and middle school. The chapter also considers the degree of urbanity, the level of development of the country, the training content, and the duration of the training for obtaining differentiated results. The chapter shows that from 21 studies with 47 estimates, 51% of the quantitative results show a positive and significant impact on student performance. This percentage is significantly greater than the 15% of results that are negative and significant. Additionally, the analysis found that some characteristics of the training, like the subject (Math or Reading Comprehension), the length of the training, and the focus (content-based or pedagogical), can increase the probability of having positive impacts. When previous findings are compared to the analysis, which only considered a strict sample that fulfills all the inclusion criteria (12 studies – 27 estimates), the results remain similar. This confirms that there are different results depending on the programs' characteristics.

For example, content-based programs, or the ones that focus on enhancing teachers' knowledge of the subject, have better returns than pedagogical programs, or ones that are more focused on improving teachers' pedagogical skills. This result can be explained by the fact that content-based programs provide teachers with general guidance on the curriculum and training for a specific subject. The pedagogical programs, in contrast, present more general content about the learning process within the classroom. However, when combining the subject and the focus of the intervention, content-based programs have a more significant impact on Math performance, while pedagogical programs have greater impacts on Reading Comprehension students' results. The reason could be the characteristics of the content-based programs for Math: these programs provide teachers with specific information on how to teach the mathematical content to their students, which has a direct impact on students' learning process.

Furthermore, more effective results are obtained when programs are sustained for a longer period of time (more than 60 hours). This result can be explained by the authors of the educational and labor sector studies referenced in the

thesis as well as by the results of the Peruvian case. Due to a longer exposure to the intervention, changes in teaching behavior are generated; according to the logical model of professional development training based on Yoon et al. (2008). These changes have a significant impact on student outcomes. Nevertheless, as none of the papers chosen for the systematic review show an optimal training duration, the question of the optimal duration of an in-service program is answered with the results of Chapter 4, when the Peruvian program's optimal duration is discussed.

Finally, contrary to what we were expecting, we discover in the systematic literature review that other variables, such as the location of professional development interventions and the development level of the country, are not related to the impact on student performance. This general analysis is expanded in Chapter 3, which studies the differential impacts of the Peruvian intervention when it is carried out in an urban versus a rural environment.

The case study used in this thesis is the Peruvian teachers' in-service program. In Chapter 3, the impact of the program on students' results in both full-grade (urban) and multigrade (rural) schools is studied using a quasi-experimental propensity score matching (PSM) design to control for differences in program enrollment like rurality, poverty, nutrition level, etc. With a dataset of 4,389 treated and 25,002 control schools for 2014, results show that the program has a positive and significant impact on Math and Reading Comprehension. For Reading Comprehension, the effect size is 0.148 for multigrade schools and 0.233 for full-grade schools. For Math, results are lower but also positive, with an effect size of 0.083 for multigrade schools and 0.137 for full-grade schools. Impacts on both subjects are lower than the impacts obtained in the international literature (around 0.4 standard deviation). The difference could be due to the characteristics of the Peruvian intervention: This intervention is mainly focused on multigrade rural schools, where students of different ages are located in the same classroom, which reduces the effect of peers and hinders the effective time of teachers for each student.

In Chapter 4, we studied the optimal length of the Peruvian program considering that, according to the official protocol, the program is supposed to train teachers for three consecutive years without any theoretical justification for this timeframe. Using propensity score matching with a difference-in-differences

analysis and data for the 2014-2016 period, this thesis estimates how incremental the impact of the program is when it last more than one year. The results show that the second year of intervention shows positive but decreasing effects on Math and Reading Comprehension for both full-grade and multigrade schools, while the third-year shows no effect on student achievement. The shortage of data for the third year of intervention led us to perform a robustness analysis dropping the extreme values of the sample (first and last quintiles), which confirms the results: the optimal length is two years of in-service training. The reasons for these decreasing results can be found in the characteristics of the training: The training is a repetitive process, wherein the second and third year of intervention only reinforce the knowledge of the previous year without any innovative component.

In Chapter 5, to obtain an optimal design of the intervention, we examine the contribution of each of the program components (classroom visits, micro workshops, and update workshops) on student achievement. As in the previous chapter, propensity score matching together with a difference-in-differences analysis is used to estimate the differential impacts of each program component for 2015. The results show that the two components closely related to the definition of "on-the-job training" (i.e. classroom visits and micro workshops) have a strong relationship with student achievement due to the way the coach interacts with the teachers: namely by discussing what happens inside the classroom. In contrast, update workshops have an almost null effect on student results, because this component tends to be carried out outside the work environment and, in general, is not related to the reality of teachers. The robustness analysis confirms the previous results.

Furthermore, the analysis found that the components like classroom visits and micro workshops, when grouped, have a greater impact than when each of them is used individually. This finding generates the recommendation that the first two components should be used together.

Chapter 5 also obtains an optimal design of the intervention using cost information of each component of the program to perform a cost-effectiveness analysis of all components. The results show that a combination of classroom visits plus micro workshops offers better results in terms of US $ per 0.1 standard deviation gain than the full intervention offers. Therefore, it is

recommendable to reconfigure the program to include only those two components.

In Chapter 6, a cost-effectiveness analysis of the whole intervention is carried out, comparing the intervention with a Peruvian program of extension of the school day as well as with eight other international educational interventions. Using information on costs and impacts for each program, it is observed that the teachers' in-service program with US$ 34 per 0.1 standard deviation gain for full-grade schools and US$ 202 per 0.1 standard deviation gain for multigrade schools, is more cost-efficient than the local intervention and is among the most cost-efficient interventions compared to the international ones.

Overall, the conclusion is that the Peruvian teachers' in-service program is having a positive impact on second graders' results in Math and Reading Comprehension in multigrade and full-grade schools and that those impacts can be improved and sustained if some changes are made in the duration and structure of the program. The principal policy recommendations derived from the study are thus to continue investing in the program as it is indeed generating a positive impact on students who live in vulnerable conditions. But this continuity of coaching for each teacher should only be carried out for two years since the study has discovered that the second year of intervention has positive but diminishing results, making a third year of intervention inadvisable. Additionally, as only the classroom visits and micro workshop components have relevant impacts on students, it is also advisable to eliminate the update workshops component and reallocate resources to strengthen the first two, since they focus on discussion and feedback between colleagues about their experiences in the classroom.

Finally, as this focalized program is cost-effective in comparison to other educational interventions, it can be used by other governments of Latin America and developing countries that are also seeking to improve the educational results of students located in rural environments. These students are usually the most vulnerable ones.

Addendum on valorization

This addendum presents the value created by this doctorate thesis in terms of the relevance of the results for non-academic audiences. It is organized in two parts describing the relevance of the findings and the target groups that will be benefited from those results; and the methodological innovations applied for the analysis.

The objective of this thesis was to quantify the impact of the Peruvian teachers in-service training on student achievement. Considering it is an ongoing intervention focused on vulnerable students mainly located in rural areas, the evaluation acquires greater importance because its results can be used to make changes to the training in order to improve its future results.

To the best of my knowledge, after the last two systematic literature reviews conducted by Kennedy (1998) and Yoon et al. (2008) in developed economies until 2005, this thesis fills a gap in the information about the effectiveness of in-service teacher training programs delivered for the last 10 years (until 2015) in developed and developing economies. Therefore, results presented in Chapter 2 of those programs characteristics that increase the probability of having positive results such as focusing on Math rather than Reading Comprehension, and using intense and sustained (more than 60 hours) content-based programs, can be used to improve the educational policies intended to reduce education inequalities in developing countries.

The empirical Chapters 3 to 5 presented in this thesis develop the questions about the impacts of the Peruvian teachers in-service training on student performance. Complementing the latter, an analysis of the optimal duration of this training, as well as the differentiated impacts of each of its three components, is developed to improve the impact of the program.

The main conclusion of this thesis is that the in-service training program do have a positive effect on student performance, both for Math and Reading comprehension for multigrade (rural) and full-grade (urban) schools; and these positive effects are sustained every year with small changes during all the period of analysis (2014-2016). The results of this thesis are closely followed by the Ministry of Education in Peru, mainly the Directorate of Primary Education and

the Directorate of Educational Services in the Rural Area, because both as responsible of the program, are in a process of evaluating the training and its components. The policy recommendation derived from this finding is to continue delivering the program, and if possible to expand it to the 63% of targeted primary schools that nowadays are not receiving the intervention because of budget constraints. A second recommendation, derived from Chapter 4, must be to reduce the duration of the program because the positive impacts disappear after two years of intervention. Thesis found decreasing marginal impact on second graders, and following the economic literature about human capital this is understood to be due to the lack of changes in the contents that are worked with each teacher, since each year the training delivers repetitive methodologies and contents. Additionally, in Chapter 5 this thesis proposes to change the structure of the program, since only the components related with a direct contact between coach and teachers (classroom visits and micro workshops) have a strong relationship with student achievement.

But in spite of these positive results, to recommend its scaling up to a national level, the question about the cost-effectiveness of public policy remained pending. This is where a second public body interested in the results of the thesis intervenes: The Ministry of Economics and Finance. Since this program is part of the Strategic Program for ¨Learning improvement of regular basic education students¨ (PELA), the interest of the Ministry is also to know the cost-effectiveness ratio of the intervention by an independent evaluation. For this reason, a special effort in Chapter 6 was made to locate the costs of the intervention at the rural level and thus be able to compare this ratio with the ones obtained from similar programs in Peru and internationally. It was found that the program is more cost-efficient than a comparable local intervention aimed at increasing lecture hours for high school students, and also with respect to other public educational interventions in the United States, such as programs aimed at increasing the salaries of the teachers, the use of payment coupons (vouchers) and the reduction of class size.

Last but not least, the results mentioned above are relevant for subnational governments that want to assess the impact of the intervention on their students. As the thesis analyzes the impact at the national level, it was not possible to discriminate the results by regions. What has happened in the last months of 2018 is that regional governments with different levels of poverty,

from the less poor such as Tacna and Arequipa, to those with the worst indices of development such as Huancavelica and Apurimac, are interested in knowing their results so that the intervention can generate greater results in its students. A clue about the need to obtain this differentiated analysis is found in the study carried out by Majerowicz y Montero (2018), which confirms the positive results of the training at the national level, but offers information on the differences in results among students with low previous performance (generally located in poor areas) and those with higher previous performance. (located in places with better social indicators).

With respect to the innovation capacity of this research, first this thesis uses information about student performance instead of focusing on the learning of the teacher receiving the training. This thesis explains the reasons for establishing a direct connection between the teachers' learning and better teaching techniques, which finally allows us to focus on the results of the national tests of student achievement, which is the final product expected by the government. Thus, when analyzing student results, we approach measuring the impact on human capital, which is directly related to the future of student income, as argued by Blau and Kahn (2005) and Checchi and van de Werfhorst (2017).

Secondly, the thesis uses a comparison methodology to compare the treated schools with teachers who receive training with those who have not been intervened. Using a Propensity Score Matching (PSM) combined in some chapters with a difference-in-differences analysis, this thesis finds the impacts of the program once the effects of other observable factors that might have biased the final results were "cleaned up". From the review of the literature, only four of the 21 papers selected in Chapter 2 use these matching techniques to find the impacts on students, which means that the present study adds evidence for the use of this methodology to find the impacts of similar educational programs.

Curriculum Vitae

Camilo Carrillo Purin was born on September 23, 1976 in Lima, Peru. He completed his Bachelor Degree in Economics at Universidad del Pacífico in December 2001. After graduating, he began working for the Peruvian government evaluating public investment projects in the Ministry of Economics and Finance, where for 4 years he was an investment analyst, a position that allowed him to travel throughout the country analyzing investment projects in the education and health sectors. From 2006 to 2007, he studied at Pompeu Fabra University (Barcelona) to obtain a Master's Degree in Economics and from 2007 to 2008 at Tilburg University (The Netherlands) to complete a Master's Degree in International Economics and Finance. Upon returning to Peru in 2009, he worked jointly with the IFC-World Bank and the Research Center of Universidad del Pacífico in the development of health investment projects for local governments in areas of poverty in Ayacucho-Peru.

From 2012 to 2014 he worked as an advisor to the Minister of Agriculture, which allowed him to occupy a position in the Board of Directors of AGROBANCO (the Agrarian Bank of Peru) and learn about the financial analysis of agricultural productive projects. Along with the beginning of the doctorate in Maastricht, he worked in the period 2014-2016 in the Ministry of Education as General Director of Infrastructure, being in charge of the development of the National Plan of Infrastructure in education; and later, from 2016 to 2018, as General Director of Private Investment of the Ministry of Economics and Finance, responsible for the development of the Public-Private Partnerships (PPP) program in the country. Since July 2018, he works as a consultant for the World Bank in PPP projects in the education sector, specializing in countries in Latin America and the Caribbean.

His research interests are focused on investment in public and public-private infrastructure, as well as in the development and evaluation of public policies.

And about his passion: it is and always has been the modern and contemporary history.

UNU-MERIT/MGSoG Dissertation Series

2019

Hugo Confraria
Developing scientific capacity in the Global South
UNU-MERIT/MGSoG Dissertation Series № 224

Alison Cathles
Educational Pathways and Skills: Past, Present, and Future
UNU-MERIT/MGSoG Dissertation Series № 223

Ibrahima Sory Kaba
Aggregate Fluctuations and Development: Essays on Macroeconomic Volatility and Economic Growth
UNU-MERIT/MGSoG Dissertation Series № 222

Charlotte Keijser
Firm Participation, Learning and Innovation in Heterogenous Value Chains of IT-enabled Services
UNU-MERIT/MGSoG Dissertation Series № 221

Salih Çevikarslan
Innovation Strategies and Their Implications for Technological Change and Market Outcomes: An Evolutionary Multi-Agent Based Modelling Approach
UNU-MERIT/MGSoG Dissertation Series № 220

Wondimagegn Mesfin Tesfaye
Essays on the Impacts of Climate-Smart Agricultural Innovations on Household Welfare
UNU-MERIT/MGSoG Dissertation Series № 219

Tatevik Poghosyan
How Board Networks Affect Firm Performance and Innovation Incentives in Transition Economies: The Case of Armenia
UNU-MERIT/MGSoG Dissertation Series № 218

Arip Muttaqien
Essays on Inequality and Polirization: Empirical Studies in Developing Asia
UNU-MERIT/MGSoG Dissertation Series № 217

2018

Katrin Marchand
Essays on Forced Migration and Labour Market Participation in Developing Countries
UNU-MERIT/MGSoG Dissertation Series № 216

Ortrun Merkle
The Myth of Gender Neutral Power: Corruption and Gender Norms
UNU-MERIT/MGSoG Dissertation Series № 215

Biljana Meshkovska
Life after Trafficking: (re)integration processes of women that have been trafficked for the purpose of sexual exploitation in Europe
UNU-MERIT/MGSoG Dissertation Series № 214

Vincenzo Vinci
The Relevance of Institutions and People's Preferences for Social Protection
UNU-MERIT/MGSoG Dissertation Series № 213

Silke Heuser
The Effectiveness of Environmental Policies on Reducing Deforestation in the Brazilian Amazon
UNU-MERIT/MGSoG Dissertation Series № 212

Jennifer Waidler
Social Assistance and Remittances and Their Role in the Fight Against Poverty
UNU-MERIT/MGSoG Dissertation Series № 211

Choolwe Muzyamba
The role of community mobilization in the promotion of maternal health of women living with HIV in Zambia
UNU-MERIT/MGSoG Dissertation Series № 210

Juan Carlos A. Castillo Sánchez
Assessing the Role of the Export Sector in Mexican Economic Development,1965-2014
UNU-MERIT/MGSoG Dissertation Series № 209

Tareq Abuelhaj
Food Security Policy Impact Analysis: The Econometrics of Cash and Food Assistance Cost Effectiveness
UNU-MERIT/MGSoG Dissertation Series № 208

Marta Férnandez de Arroyabe Arranz
Essays on M&AS and Innovation
UNU-MERIT/MGSoG Dissertation Series № 207

Clotilde Mahé
Essays on Migration and Occupational Choice
UNU-MERIT/MGSoG Dissertation Series № 206

Simone Sasso
Talent on the move. Essays on Human Capital, Graduate Mobility and Economic Development
UNU-MERIT/MGSoG Dissertation Series № 205

Khaled Walid Rajab
Strategic Planning under Fragility
UNU-MERIT/MGSoG Dissertation Series № 204

Mutinta Hambayi Nseluke
A Tall Order: Improving Child Linear Growth
UNU-MERIT/MGSoG Dissertation Series № 203

Elvis Korku Avenyo
Innovations and Firm Performance in sub-Saharan Africa: Empirical Analyses
UNU-MERIT/MGSoG Dissertation Series № 202

Ni Zhen
Employment Dynamics, Firm Performance and Innovation Persistence in the Context of Differentiated Innovation Types: Evidence from Luxembourg
UNU-MERIT/MGSoG Dissertation Series № 201

Caroline Wehner
Too Scared to Achieve: The Relation Between Neuroticism, Conscientiousness and Socioeconomic Outcomes
UNU-MERIT/MGSoG Dissertation Series № 200

Stefania Innocenti
On Institutional Persistence
UNU-MERIT/MGSoG Dissertation Series № 199

Hassen Abda Wako
Economic Globalization, Institutions and Development: Essays on Aid, Foreign Direct Investment and Trade
UNU-MERIT/MGSoG Dissertation Series № 198

2017

Hans-Erik Edsand
Winds of Change
UNU-MERIT/MGSoG Dissertation Series № 197

Ana Patricia Silva Vara
Redressing the Gender Gap
UNU-MERIT/MGSoG Dissertation Series № 196

Andrés Iván Mideros Mora
Essays on the Economic Effects of Non-contributory Social Protection
UNU-MERIT/MGSoG Dissertation Series № 195

Tobias Broich
New Actors in the Global Economy
UNU-MERIT/MGSoG Dissertation Series № 194

Bernard Nikaj
From No-government to E-government
UNU-MERIT/MGSoG Dissertation Series № 193

Ali Safarnejad
Prioritizing the HIV Response
UNU-MERIT/MGSoG Dissertation Series № 192

Clovis Freire
Diversification and Structural Economic Dynamics
UNU-MERIT/MGSoG Dissertation Series № 191

Michael Verba
Innovation and Knowledge Dynamics: Essays on the Knowledge Economy
UNU-MERIT/MGSoG Dissertation Series № 190

Pui Hang Wong
The Hearts and Minds in Conflict and Peace: The Economics of Counterinsurgency and the Psychology of Reconstruction
UNU-MERIT/MGSoG Dissertation Series № 189

Brenda Yamba
Schooling Despite All Odds: Evidence from Lesotho on Female Child Carers who Stayed in School
UNU-MERIT/MGSoG Dissertation Series № 188

Sheng Zhong
Moving towards An Energy Efficient Future: Essays on Energy Efficiency, Technology and Development
UNU-MERIT/MGSoG Dissertation Series № 187

Julieta Marotta
Access to Justice and Legal Empowerment of Victims of Domestic Violence through Legal Organizations in the City of Buenos Aires: A Qualitative Empirical Legal Study
UNU-MERIT/MGSoG Dissertation Series, № 186

Andrea Franco-Correa
On the Measurement of Multidimensional Poverty as a Policy Tool: Empirical Applications to Chile, Colombia, Ecuador and Peru
UNU-MERIT/MGSoG Dissertation Series, № 185

2016

Yesuf Awel
Insurance for Growth: Empirical Essays on Insurance Demand and Impacts in Africa
UNU-MERIT Dissertation Series, № 108

Tigist Mekonnen Melesse
Grow More Food using Fewer Resources: Agricultural Technology Adoption and Innovation Practices for Inclusive and Sustainable Development
UNU-MERIT Dissertation Series, № 107

Eleni Yitbarek
Getting Ahead or left Behind? Essays on Poverty Dynamics and Social Mobility in Africa
UNU-MERIT Dissertation Series, № 106

Thuy Dieu Nguyen
Firm-Level Theory and Evidence of Corruption
UNU-MERIT Dissertation Series, № 105

Raquel Tsukada Lehman
Essays on Household Production with Labor-Saving Technology
UNU-MERIT Dissertation Series, № 104

Eva Barteková
Multi-Problem Challenges for a Renewable Future: Empirical Studies on Competitive Disadvantages from Electricity Price Differentials and Mineral Supply Risk in an Open Economy
UNU-MERIT Dissertation Series, № 103

Jocelyn Olivari
Entrepreneurial Traits and Innovation: Evidence from Chile
UNU-MERIT Dissertation Series, № 102

Muhammad Shafique
Essays on the role of knowledge, R&D, and Technology-based Firms in the Evolution of Socio-techno-economic System
UNU-MERIT Dissertation Series, № 101

Serdar Türkeli
Governance of Innovation Policy: Empirical Studies on Applied Political Economy by Multi-Methods Analysis
UNU-MERIT Dissertation Series, № 100

Ayokunu Adedokun
Pathways to Sustainable Peace building in Divided Societies: Lessons and Experiences from Mozambique
MGSoG Dissertation Series, № 75

Luiz Rothier Bautzer
Organizing Concurrent Engineering through ICT Platforms Blueprinting Product Lifecycle Management Platforms across Disciplinary Agencies
MGSoG Dissertation Series, № 74

Natalia Popova
Migration in the Periphery of the European Union: Determinants of Successful and Sustainable Labour Market Integration of Return Migrants in Albania, Egypt, Moldova and Tunisia
MGSoG Dissertations Series, № 73

Richard A. Martina
Uncertainty and Resource Constraint in the Small Island Developing States: Essays in Entrepreneurial Cognition
MGSoG Dissertations Series, № 72

Cécile Cherrier
The Expansion of Basic Social Protection in Low-income Countries: An Analysis of Foreign Aid Actors' Role in the Emergence of Social Transfers in Sub-Saharan Africa
MGSoG Dissertations series, № 71

Paul Caldron
The Tacit Bargain in Short-Term Medical Missions: Why U.S. physicians go and what it costs
MGSoG Dissertation Series, № 70

Mahmut Kobal
Customs & Excellence: A Comparative Approach on Administrative and Regulatory Compliance Perspectives of the EU-Turkey Customs Union
MGSoG Dissertation Series, № 69

Craig Loschmann
Essays on Conflict-related Migration and Development in the Case of Afghanistan
MGSoG Dissertations Series, № 68

Andrea Milan
Rural Livelihoods, Location and Vulnerable Environments: Approaches to Migration in Mountain areas of Latin America
MGSoG Dissertation Series, № 67

Farida Lada
On Guarding the Welfare of Clinical Trial Subjects While Promoting Novel Drug Innovation
A Game Theoretical Approach
MGSoG Dissertation Series, № 66

2015

Hibret Belete Maemir
Dissecting Aggregate Productivity: International Integration and Growth with Heterogeneous Firms
UNU-MERIT Dissertation Series, № 96

Giorgio Triulzi
Looking for the Right Path: Technology Dynamics, Inventive Strategies and Catching-up in the Semiconductor Industry
UNU-MERIT Dissertation Series, № 95

Abdul Baseer Qazi
Knowledge flows and networks in the ICT sector: The case of Pakistan
UNU-MERIT Dissertation Series, № 94

Ajay Thutupalli
Technology Paradigm Shifts in Agriculture: Drivers of Sustainability and Catch up
UNU-MERIT Dissertation Series, № 93

Eduardo Urias
Improving access to HIV/AIDS treatment in Brazil: When are Compulsory Licenses effective in Price Negotiations?
UNU-MERIT Dissertation Series, № 92

Francesca Guadagno
Why have so few Countries Industrialised?
UNU-MERIT Dissertation Series, № 91

Daniel Opolot
The Evolution of Beliefs and Strategic Behaviour
UNU-MERIT Dissertation Series, № 90

Alejandro Lavopa
Structural Transformation and Economic Development: Can Development Traps be Avoided
UNU-MERIT Dissertation Series, № 89

Jinjin Zhao
Urban water management reform: The Case of China
UNU-MERIT Dissertation Series, № 88

Simona Vezzoli
Borders, Independence and Post-colonial Ties: the Role of the State in Caribbean Migration
MGSoG Dissertation Series, № 65

Silvia Consuelo Gómez Soler
Civil Conflict and Education: How Does Exposure to Civil Conflict Affect Human Capital Accumulation? Evidence from Standardized Exit Exams in Colombia
MGSoG Dissertation Series, № 64

Paula Nagler
Occupational Choice in the Developing World
MGSoG Dissertation Series, № 63

Jasmin Kientzel
Determinants of Professional Commitment to Environmental Sustainability
MGSoG Dissertation Series, № 62

Mehmet Güney Celbiş
Regional Policies: Convergence, Trade, and the Allocation of Public Capital
MGSoG Dissertation Series, № 61

Florian Henning
Living Up to Standard: Interoperability Governance and Standards Adoption in Government Information Networks
MGSoG Dissertation Series, № 60

Niels P. Groen
The Never-Ending Project Understanding E-Government Project Escalation
MGSoG Dissertation Series, № 59

Derek Copp
Teacher-Based Reactivity to Provincial Large-scale Assessment in Canada
MGSoG Dissertation Series, № 58

Michaella Vanore
Family-Member Migration and the Psychosocial Health Outcomes of Children in Moldova and Georgia
MGSoG Dissertation Series, № 57

Sonja Fransen
The Economic and Social Effects of Remittances and Return Migration in Conflict-Affected Areas: The Case of Burundi
MGSoG Dissertation Series, № 56

Ibrahim Khalil Conteh
The Impact of Floods on Primary School Education in Zambia
MGSoG Dissertation Series, № 55

Richard Bluhm
Growth Dynamics and Development Essays in Applied Econometrics and Political Economy
MGSoG Dissertation Series, № 54

Nevena P. Zhelyazkova
Work-Family Reconciliation and Use of Parental Leave in Luxembourg: Empirical Analysis of Administrative Records
MGSoG Dissertation Series, № 53

2014

Dirk Crass
The Impact of Brands on Innovation and Firm Performance: Empirical Evidence from Germany
UNU-MERIT Dissertation Series, № 87

Samyukta Bhupatiraju
The Geographic Dimensions of Growth and Development
UNU-MERIT Dissertation Series, № 86

François Lafond
TheEvolution of Knowledge Systems
UNU-MERIT Dissertation Series, № 85

Annalisa Primi
Promoting Innovation in Latin America: What Countries Have Learned (and What They Have Not) in Designing and Implementing Innovation and Intellectual Property Policies
UNU-MERIT Dissertation Series, № 84

Fatoumata Lamarana Diallo
Evaluation of Meal and Deworming Programs for Primary Schools in Rural Senegal
UNU-MERIT Dissertation Series, № 83

Sachin Kumar Badkas
Metachoice and Metadata: Innovating with Environmental Policy Analysis in Europe
MGSoG Dissertation Series, № 52

Irina S. Burlacu
An Evaluation of Tax-Benefit Systems Impact on the Welfare of Frontier Worker:
The Case of Luxembourg and Belgium
MGSoG Dissertation Series, № 51

Özge Bilgili
Simultaneity in Transnational Migration Research: Links Between Migrants' Host and Home Country Orientation
MGSoG Dissertation Series, № 50

Yulia Privalova Krieger
Reshaping the Big Agenda: Transnational Politics and Domestic ResistanceFinancial crisis and social protection reform in Bosnia and Herzegovina
MGSoG Dissertation Series, № 49

Marieke van Houte
Moving Back or Moving Forward? Return migration after Conflict
MGSoG Dissertation Series, № 48

Oxana Slobozhan
Global Governance in the Management of Natural Resources: The Case of the Extractive Industries Transparency Initiative (EITI)
MGSoG Dissertation Series, № 47

Luis Bernardo Mejia Guinand
The Changing Role of the Central Planning Offices in Latin America: A Comparative Historical Analysis Perspective (1950-2013)
MGSoG Dissertation Series, № 46

Cheng Boon Ong
Ethnic Segregation in Housing, Schools and Neighbourhoods in the Netherlands
MGSoG Dissertation Series, № 45

Luciana V. Cingolani
Bureaucracies for Development: Oxymoron or Reality? Studies on State Capacity in Challenging Governance Contexts
MGSoG Dissertation Series, № 44

Carlos Cadena Gaitán
Green Politics in Latin American Cities - Sustainable Transport Agendas
MGSoG Dissertation Series, № 43

Katie Kuschminder
Female Return Migration and Reintegration Strategies in Ethiopia
MGSoG Dissertation Series, № 42

Metka Hercog
Highly-Skilled Migration and New Destination Countries
MGSoG Dissertation Series, № 41

Margaret Agaba Rugadya
Can Remittances Influence the Tenure and Quality of Housing in Uganda?
MGSoG Dissertation Series, № 40

Ilire Agimi
New Governance Under Limited Statehood: The Case of Local Government Reform in Kosovo
MGSoG Dissertation Series, № 39

2013

Anant Kamath
Information Sharing through Informal Interaction in Low-Tech Clusters
UNU-MERIT Dissertation Series, № 82

Flavia Pereira de Carvalho
What we talk about when we talk about Brazilian Multinationals: An Investigation on Brazilian FDI, Economic Structure, Innovation and the Relationship between them
UNU-MERIT Dissertation Series, № 81

Jun Hou
Complementarity in Innovation and Development: A Cross-country Comparison
UNU-MERIT Dissertation Series, № 80

Rufin Baghana
Impacts of Government Incentives to R&D, Innovation and Productivity: A Microeconometric Analysis of the Québec Case
UNU-MERIT Dissertation Series, № 79

Lilia I. Stubrin
High-Tech Activities in Emerging Countries: A Network perspective on the Argentinean Biotech Activity
UNU-MERIT/MGSoG Dissertation Series, № 78

Kristine Farla
Empirical Studies on Institutions, Policies and Economic Development
MGSoG Dissertation Series, № 38

Marina Petrovic
Social Assistance and Activation in the Pursuit of Happiness: Shedding New Light on Old Policy Solutions to Social Exclusion
MGSoG Dissertation Series, № 37

Laura Torvinen
Assessing Governance Assessments: The Case of Mozambique: Governance Assessments in the Context of Aid Effectiveness Discourse
MGSoG Dissertation Series, № 36

Biniam Egu Bedasso
Institutional Change in the Long Shadow of Elite: Essays on Institutions, Human Capital and Ethnicity in Developing Countries
MGSoG Dissertation Series, № 35

Sepideh Yousefzadeh Faal Deghati
Childhoods Embargoed: Constructing and Reconstructing Multidimensional Child Poverty in Iran 1984-2009
MGSoG Dissertation Series, № 34

Robert Bauchmüller
Investing in Early Childhood Care and Education: The Impact of Quality on Inequality
MGSoG Dissertation Series, № 33

Martin Rehm
Unified Yet Separated: Empirical Study on the Impact of Hierarchical Positions within Communities of Learning
MGSoG Dissertation Series, № 32

2012

Abdul Waheed
Innovation Determinants and Innovation as a Determinant: Evidence from Developing Countries
UNU-MERIT Dissertation Series, № 77

Bilal Mirza
Energy Poverty and Rural Energy Markets in Pakistan
UNU-MERIT Dissertation Series, № 76

Benjamin Engelstätter
Enterprise Software and Video Games: An Empirical Analysis
UNU-MERIT Dissertation Series, № 75

Fulvia Farinelli
Natural Resources, Innovation and Export Growth: The Wine Industry in Chili and Argentina
UNU-MERIT Dissertation Series

Rodolfo Lauterbach
Innovation in Manufacturing: From Product Variety and Labor Productivity Growth to Economic Development in Chile
UNU-MERIT Dissertation Series

Kirsten Wiebe
Quantitative Assessment of Sustainable Development and Growth in Sub-Saharan Africa
UNU-MERIT Dissertation Series, № 74

Julio Miguel Rosa
Organizational Strategies, Firms' Performance and Spatial Spillovers: The Canadian Case in Research and Development.
UNU-MERIT Dissertation Series, № 73

Johannes Wilhelmus Marie Boels
Joseph Schumpeter, Honderd Jaar Economische Ontwikkeling: Een Historisch-theoretische Beschouwing.
UNU-MERIT Dissertation Series

Dorcas Mbuvi
Utility Reforms and Performance of the Urban Water Sector in Africa
MGSoG Dissertation Series, № 31

Lina Salanauskaite
Distributional Impacts of Public Policies: Essays in Ex-Ante and Ex-Post Evaluation
MGSoG Dissertation Series, № 30

Esther Schüring
To Condition or not – is that the Question?
An Analysis of the Effectiveness of Ex-Ante and Ex-Post Conditionality in Social Cash Transfer Programs
MGSoG Dissertation Series, № 29

Joe Abah
Strong Organisations in Weak States: Atypical Public Sector Performance in Dysfunctional Environments
MGSoG Dissertation Series, № 28

Zina Samih Nimeh
Social Citizenship Rights: Inequality and Exclusion
MGSoG Dissertation Series, № 27

2011

Daniel Vertesy
Interrupted Innovation: Emerging Economies in the Structure of the Global Aerospace Industry
UNU-MERIT Dissertation Series, № 72

Tina Saebi
Successfully Managing Alliance Portfolios: AnAlliance Capability View
UNU-MERIT Dissertation Series, № 71

Nora Engel
Tuberculosis in India: A Case of Innovation and Control
UNU-MERIT/MGSoG Dissertation Series, № 70

Evans Mupela
Connectivity and growth in Sub-Saharan Africa: The Role of Communication Satellites
UNU-MERIT Dissertation Series, № 69

Nantawan Kwanjai
Cross Cultural Intelligence amid Intricate Cultural Webs: A Tale of the UnDutchables in the Land of 1002 Smiles
UNU-MERIT Dissertation Series, № 68

Lina Sonne
Innovation in Finance to Finance Innovation: Supporting Pro-poor Entrepreneur-based Innovation
UNU-MERIT Dissertation Series, № 67

Lenka Eisenhamerová
Legitimacy of 'Humanitarian Military Intervention'
MGSoG Dissertation Series, № 26

Sonila Tomini
Informal Payments for Health Care Services in Albania
MGSoG Dissertation Series, № 25

Jinjing Li
Dynamic Microsimulation in Public Policy Evaluation
MGSoG Dissertation Series, № 24

Aziz Atamanov
Rural Nonfarm Employment and International Migration as Alternatives to Agricultural Employment: The Case of Kyrgyzstan
MGSoG Dissertation Series, № 23

Frieda Vandeninden
Poverty Alleviation: Aid and Social Pensions
MGSoG Dissertation Series, № 22

Juliana Nyasha Tirivayi
The Welfare Effects of Integrating AIDS Treatment with Food Transfers: Evidence from Zambia
MGSoG Dissertation Series, № 21

Agnieska Ewa Sowa
Who's Left Behind? Social Dimensions of Health Transition and Utilization of Medical Care in Poland
MGSoG Dissertation Series, № 20

Emmanaouil Sfakianakis
The Role of Private Actors in the Provision of Public Goods with Applications to Infrastructure and Financial Stability
MGSoG Dissertation Series, № 19

Siu Hing Lo
White Collars Green Sleeves: An Inter-organizational Comparison of Determinants of Energy-Related Behaviors among Office Workers
MGSoG Dissertation Series, № 18

Treena Wu
Constraints to Human Capital Investment in Developing Countries: Using the Asian Financial Crisis in Indonesia as a Natural Experiment
MGSoG Dissertation Series, № 17

Henry Espinoza Peña
Impact Evaluation of a Job-Training Programme for Disadvantaged Youths: The Case of Projoven
MGSoG Dissertation Series, № 16

2010

Fernando Santiago
Human Resources Management Practices and Learning for Innovation in Developing Countries: Pharmaceutical Firms in Mexico
UNU-MERIT Dissertation Series, № 66

Zakaria Babutsidze
Essays on Economies with Heterogeneous Interacting Consumers
UNU-MERIT Dissertation Series, № 65

Bertha Vallejo
Learning and Innovation Under Changing Market Conditions: The Auto Parts Industry in Mexico
UNU-MERIT Dissertation Series, № 64

Donatus Ayitey
Technical Change, Competitiveness and Poverty Reduction: A Study of the Ghanaian Apparel Industry
UNU-MERIT Dissertation Series, № 63

Sergey Filippov
Multinational Subsidiary Evolution: Corporate Change in New EU Member States
UNU-MERIT Dissertation Series, № 62

Asel Doranova
Technology Transfer and Learning under the Kyoto Regime: Exploring the Technological Impact of CDM Projects in Developing Countries
UNU-MERIT Dissertation Series, № 61

Florian Tomini
Between Family and Friend: Understanding the Interdependency of Private Transfers
MGSoG Dissertation Series, № 15

Michał Polalowski
The Institutional Transformation of Social Policy in East Central Europe: Poland and Hungary in Comparative and Historical Perspective
MGSoG Dissertation Series, № 14

Maha Ahmed
Defining, Measuring and Addressing Vulnerability: The Case of Post Conflict Environments
MGSoG Dissertation Series, № 13

Pascal Beckers
Local Space and Economic Success: The Role of Spatial Segregation of Migrants in the Netherlands
MGSoG Dissertation Series, № 12

Victor Cebotari
Conflicting Demands in Ethnically Diverse Societies: Ethno political Contention and Identity Values in Europe
MGSoG Dissertation Series, № 11

Dennis Gyllensporre
Competing and Complementary Perspectives on the EU as a Crisis Management Actor: An Examination of the Common Security and Defence Policy through the Lenses of Idealism and Realism
MGSoG Dissertation Series, № 10

Judit Vall Castello
Business Cycle and Policy Effects on Labour Market Transitions of Older and Disabled Workers in Spain
MGSoG Dissertation Series, № 9

Keetie Roelen
False Positives or Hidden Dimensions: The Definition and Measurement of Child Poverty
MGSoG Dissertation Series, № 8

Denisa Maria Sologon
Earning Dynamics in Europe
MGSoG Dissertation Series, № 7

Melissa Siegel
Money and Mobility: Migration and Remittances
MGSoG Dissertation Series, № 6

Jessica S. Hagen-Zanker
Modest Expectations: Causes and Effects of Migration on Migrant Households inSource Countries
MGSoG Dissertation Series, № 5

2009

Alexis Habiyaremye
From Primary Commodity Dependence to Diversification and Growth: Absorptive Capacity and Technological Catch Up in Botswana and Mauritius.
UNU-MERIT Dissertation Series, № 60

Yoseph Getachew
The Role of Public Capital in Economic Development
UNU-MERIT Dissertation Series, № 59

Sandra Leitner
Embodied Technological Change and Patterns of Investment in Austrian Manufacturing
UNU-MERIT Dissertation Series, № 58

Semih Akçomak
The Impact of Social Capital on Economic and Social Outcomes
UNU-MERIT Dissertation Series, № 57

Abraham Garcia
The Role of Demand in Technical Change
UNU-MERIT Dissertation Series, № 56

Saurabh Arora
Coherence in Socio-technical Systems: A Network Perspective on the Innovation Process
UNU-MERIT Dissertation Series, № 55

Mirtha R. Muniz Castillo
Human Development and Autonomy in Project Aid: Experiences from four bilateral projects in Nicaragua and El Salvador
MGSoG Dissertation Series, № 4

Christiane Arndt
Governance Indicators
MGSoG Dissertation Series, № 3

Britta Augsburg
Microfinance: Greater Good or Lesser Evil?
MGSoG Dissertation Series, № 2

2008

Rutger Daems
Medicines for the Developing World
UNU-MERIT Dissertation Series, № 54

Johannes Hanel
Assessing Induced Technology: Sombart's Understanding of Technical Change in the History of Economics
UNU-MERIT Dissertation Series, № 53

Rifka Weehuizen
Mental Capital: the Economic Significance of Mental Health
UNU-MERIT Dissertation Series, № 52

Danielle Cloodt
The Relationship between R&D Partnership Formation, Social Embeddedness and Innovative Performance
UNU-MERIT Dissertation Series, № 51

Sabine Fuss
Sustainable Energy Development under Uncertainty
UNU-MERIT Dissertation Series, № 50

Geranda Notten
Measuring and Managing Poverty Risks
MGSoG Dissertation Series, № 1

2007

Tobias Kronenberg
Reconciling Environmental Conservation with Economic Prosperity: The Feasibility of Double Dividends in the Short and Long Run
UNU-MERIT Dissertation Series, № 49

Viktoria Kravtsova
Assessing the Impact of Foreign Direct Investment in Transition Economies
UNU-MERIT Dissertation Series, № 48

Suhail Sultan
The Competitive Advantage of Small and Medium Sized Enterprises: The Case of Jordan's Natural Stone Industry
UNU-MERIT Dissertation Series, № 47

2006

Bulat Sanditov
Essays on Social Learning and Imitation
UNU-MERIT Dissertation Series, № 46

Mamata Parhi
Dynamics of New Technology Diffusion: A Study of the Indian Automotive Industry
UNU-MERIT Dissertation Series, № 45

Andreas Reinstaller
Social Structures and the Innovation Process: Their Role in the Demand of Firms and Consumers
UNU-MERIT Dissertation Series, № 44

Rose Kiggundu
Innovation systems and Development: The Journey of a Beleaguered Nile Perch Fishery in Uganda
UNU-MERIT Dissertation Series, № 43

Thomas Pogue
The Evolution of Research Collaboration in South African Gold Mining: 1886-1933
UNU-MERIT Dissertation Series, № 42

Geoffrey Gachino
Foreign Direct Investment, Spillovers and Innovation: The Case of Kenyan Manufacturing Industry
UNU-MERIT Dissertation Series, № 41

Önder Nomaler
Technological Change, International Trade and Growth: An Evolutionary, Multi-Agents-Based Modeling Approach
UNU-MERIT Dissertation Series, № 40

2005

Samia Satti Osman Mohamed-Nour
Change and Skill Development in the Arab Gulf Countries
UNU-MERIT Dissertation Series, № 39

Elad Harison
Intellectual Property Rights: Economics and Policy Analysis
UNU-MERIT Dissertation Series, № 38

Daniel Dalohoun
Learning to innovate: agricultural innovation and entrepreneurship: the case of Songhaï farmers in Benin
UNU-MERIT Dissertation Series, № 37

Müge Ozman
Networks, Organizations and Knowledge
UNU-MERIT Dissertation Series, № 36

Bas Straathof
Product Variety and Economic Growth: The Counteracting Effects of Scale and Idiosyncrasy
UNU-MERIT Dissertation Series, № 35

Wilfred Schoenmakers
Knowledge Flows between Multinational Companies: A Patent Data Analysis
UNU-MERIT Dissertation Series, № 34

Myriam Cloodt
Mergers and Acquisitions (M and As) in High-Tech Industries: Measuring the Post-M and A Innovative Performance of Companies
UNU-MERIT Dissertation Series, № 33

2004

Paola Criscuolo
R&D Internationalisation and Knowledge Transfer: Impact on MNEs and their Home Countries
UNU-MERIT Dissertation Series, № 32

Maarten Verkerk
Trust and Power on the Shop Floor
UNU-MERIT Dissertation Series, № 31

Gottfried Leibbrandt
Adoption, Harmonization and Succession of Network Technologies across Countries
UNU-MERIT Dissertation Series, № 30

Mark Sanders
Skill Biased Technical change: Its Origins, the Interaction with the Labour Market and Policy Implications
UNU-MERIT Dissertation Series, № 29

2003

Nadine Roijakkers
Inter-firm Cooperation in High-tech Industries: a Study of R&D Partnerships in Pharmaceutical Biotechnology
UNU-MERIT Dissertation Series, № 28

Viki Sonntag
Speed, Scale and Sustainability
UNU-MERIT Dissertation Series, № 27

Masaru Yarime
From End-of-Pipe Technology to Clean Technology
UNU-MERIT Dissertation Series, № 26

Stéphane Malo
The Combinatorial Chemistry Revolution: Sustaining a Superior Performance Position through Technological Learning
UNU-MERIT Dissertation Series, № 25

2002

Annelies Hogenbirk
Determinants of Inward Foreign Direct Investment: the Case of the Netherlands
UNU-MERIT Dissertation Series, № 24

Bastiaan Johan terWeel
The Computerization of the Labour Market
UNU-MERIT Dissertation Series

2001

John Adeoti
Technology Investment in Pollution Control in Sub-Saharan Africa: The Case of the Nigerian Manufacturing Industry
UNU-MERIT Dissertation Series, № 23

Edward Huizenga
Innovation Management: How Frontrunners Stay Ahead: An Empirical Study on Key Success Factors in the ICT sector
UNU-MERIT Dissertation Series, № 22

2000

Machiel van Dijk
Technological Change and the Dynamics of Industries: Theoretical Issues and Empirical evidence from Dutch Manufacturing
UNU-MERIT Dissertation Series, № 21

1999

Jan Cobbenhagen
Managing Innovation at the Company Level: A Study on Non-Sector-Specific Success Factors
UNU-MERIT Dissertation Series, № 20

Marjolein Caniëls
Regional Growth Differentials: The Impact of Locally Bounded Knowledge Spillovers
UNU-MERIT Dissertation Series, № 19

1998

Aldo Geuna
Resource Allocation and Knowledge production: Studies in the Economics of University Research
UNU-MERIT Dissertation Series, № 18

1996

Reinoud Joosten
Dynamics, Equilibria, and Values
UNU-MERIT Dissertation Series, № 17

Hugo Kruiniger
Investment, R&D, and the Financing Decisions of the Firm
UNU-MERIT Dissertation Series, № 16

1995

Hans van Meijl
Endogenous Technological Change: The Case of Information Technology, Theoretical Considerations and Empirical Results
UNU-MERIT Dissertation Series, № 15

René Kemp
Environmental Policy and Technical Change: A Comparison of the Technological Impact of Policy Instruments
UNU-MERIT Dissertation Series, № 14

Rohini Acharya
The Impact of New Technologies on Economic Growth and Trade: A Case Study of Biotechnology
UNU-MERIT Dissertation Series, № 13

Geert Duysters
The Evolution of Complex Industrial Systems: The Dynamics of Major IT Sectors
UNU-MERIT Dissertation Series, № 12

Marjan Groen
Technology, Work and Organisation: A Study of the Nursing Process in Intensive Care Units
UNU-MERIT Dissertation Series, № 11

1994

Huub Meijers
On the Diffusion of Technologies in a Vintage Framework: Theoretical Considerations and Empirical Results
UNU-MERIT Dissertation Series, № 10

Theon van Dijk
The Limits of Patent Protection: Essays on the Economics of Intellectual Property Rights
UNU-MERIT Dissertation Series, № 9

Hans Voordijk
Naar Integrale Logistiek in Bedrijfsketens: Ontwikkelingen in de Bouw
UNU-MERIT Dissertation Series, № 8

1993

Paul Diederen
Technological Progress in Enterprises and Diffusion of Innovation: Theoretical Reflections and Empirical Evidence
UNU-MERIT Dissertation Series, № 7

Ben Dankbaar
Economic Crisis and Institutional Change: The Crisis of Fordism from the Perspective of the Automobile Industry
UNU-MERIT Dissertation Series, № 6

Hanno Roberts
Accountability and Responsibility: The Influence of Organisation Design on Management Accounting
UNU-MERIT Dissertation Series, № 5

1992

Bart Verspagen
Uneven Growth between Interdependent Economies: An Evolutionary View on Technology Gaps, Trade and Growth
UNU-MERIT Dissertation Series, № 4

Sjoerd Romme
A Self-organization Perspective on Strategy Formation
UNU-MERIT Dissertation Series, № 3

1989

John Spangenberg
Economies of Scale, and Atmosphere in Research Organisations
UNU-MERIT Dissertation Series, № 2

1988

John Hagedoorn
Evolutionary and Heterodox Innovation Analysis: A Study of Industrial and Technological Development in Process Control and Information Technology
UNU-MERIT Dissertation Series, № 1