

Human capital development at school and work

Citation for published version (APA):

Ferreira Sequeda, M. (2019). *Human capital development at school and work*. [Doctoral Thesis, Maastricht University]. ROA. <https://doi.org/10.26481/dis.20190510mf>

Document status and date:

Published: 01/01/2019

DOI:

[10.26481/dis.20190510mf](https://doi.org/10.26481/dis.20190510mf)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Human Capital Development at School and Work

©M. Ferreira Sequeda, Maastricht, The Netherlands 2019

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing, from the author.

Published by ROA
P.O. Box 616
6200 MD Maastricht

ISBN: 978-90-5321-577-7

Printed in the Netherlands by Canon

Human Capital Development at School and Work

DISSERTATION

To obtain the degree of Doctor
at 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
Friday 10 May 2019 at 10.00 hours

by

Maria Ferreira Sequeda

Promotor

Prof. dr. A. de Grip

Co-promotor

Dr. A. Künn-Nelen

Assessment Committee

Prof. dr. T. Dohmen (Chairman)

Prof. dr. G. Brunello (University of Padova)

Prof. dr. A. Chevalier (Royal Holloway, University of London)

Dr. T. Schils

Ithaka

*As you set out for Ithaka
hope your road is a long one,
full of adventure, full of discovery.
Laistrygonians, Cyclops,
angry Poseidon – don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians, Cyclops,
wild Poseidon—you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.*

*Hope your road is a long one.
May there be many summer mornings when,
with what pleasure, what joy,
you enter harbors you're seeing for the first time;
may you stop at Phoenician trading stations
to buy fine things,
mother of pearl and coral, amber and ebony,
sensual perfume of every kind –
as many sensual perfumes as you can;
and may you visit many Egyptian cities
to learn and go on learning from their scholars.*

*Keep Ithaka always in your mind.
Arriving there is what you're destined for.
But don't hurry the journey at all.
Better if it lasts for years,
so you're old by the time you reach the island,
wealthy with all you've gained on the way,
not expecting Ithaka to make you rich.*

*Ithaka gave you the marvelous journey.
Without her you wouldn't have set out.
She has nothing left to give you now.*

*And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you'll have understood by then what these Ithakas mean.*

Constantine P. Cavafy: *Collected Poems*, translated by Edmund Keeley and Philip Sherrard (Princeton University Press, 1975).

Acknowledgements

I embarked on my first trip to Europe in 2011 with half of a plan (that included a Doctorate venture) and a sizable half-empty bag of all kinds of questions. The past years have been a tremendously exciting and challenging period of growth for me, and as I sit here today concluding my PhD dissertation the one word that comes to mind is *Gratitude*. Gratitude to life and all the people, organisations, situations, and events that have been part of this journey and that, in one way or another, have made it possible. Thank you!

I'd like to express my utmost appreciation to Andries, my supervisor, who has been a truly dedicated mentor. Thanks so much for giving me the opportunity to work on this research at ROA and for your continuous support, guidance, patience, and motivation. It's been a real privilege to be your mentee and learn from your extraordinary combination of scientific and human qualities. I'm also hugely appreciative to Annemarie for her valuable advice as my second supervisor. My sincerest gratitude to you both for not only serving as my supervisors but also for caring about my career and my life, especially during the tough moments. You inspired and encouraged my best efforts.

I'd also like to thank my co-authors Bart, Sergio, Rolf, Andries and Annemarie for the countless insightful conversations we have had. These and all other discussions with great people at several events and conferences have made clear to me that advancing our understanding of the world is just another lifelong expedition next to learning about our humanity.

I wish to extend my gratitude to all my ROA and SBE colleagues, the ones that remain and those who embarked into new adventures. Thanks for going far beyond the call of duty and for offering me your kind support that ranges from discussing my work in progress to helping me with the logistics of acquiring 'that' dataset or attending 'that' conference... Thanks for all the cups of coffee and tea, breakfasts, lunches, dinners, drinks, BBQs, carnivals, parties, *outjes* and even self-inflicting running or gym sessions that we shared as a way to connect to each other or simply scape the world. Of course, special mention to my PhD-mates Sanne, Inge, Merve, Kim, Seher, Maria, Omar, Sergio, Alexander and many many others, with whom I have shared plenty of the above moments and also those of deep anxiety or big excitement.

This dissertation wouldn't have been possible without the friendships I have been luckily surrounded of. I have been fortunate to come across many good friends, without whom life would be bleak. Special thanks to my friends in Colombia and the rest of the world with whom distance has never been a problem to find a time to Skype even if it's 2 in the morning here or there. As many

others, I have been far away from my beloved country and family for quite some time already. So, many thanks to all the international friends I have made while living in The Hague, Eindhoven, Maastricht and Utrecht: you have become my family here in the Netherlands! The list of names is so long that I will better mention how very much fond of memories and affection I'm because of you all. Some special words of gratitude to my friends who have always been a major source of support when things would get a bit discouraging: Caro, Pame, Pao, Angie, Inge, Sanne, Merve, and 'Ilustres Economistas' ... Thanks for always being there for me. And thanks, Dorrit: I definitely wouldn't have made it without you!

I would also like to extend my infinite gratitude to Susana Valdivieso, my very first mentor and academic friend: thank you for nurturing my enthusiasm for Economics and Social Sciences, for teaching me to believe in myself and dreaming things I didn't know about or thought were too far from my reach. Mil gracias Susana por haberme motivado a empezar este viaje y por su constante apoyo de múltiples maneras a lo largo de este proceso. Este logro se lo debo en gran parte a usted!

Many thanks as well to Héctor Otero and Ligia Cortés, my mentors and friends, who provided me with the courage to embark on this journey in the first place. You two are a role of true love and family happiness I look up to. Ustedes han sido muy importantes en mi camino y en esta etapa en particular. Gracias por el ánimo y energía que siempre me transmiten, no matter what.

None of this would have been possible without the love and patience of my family. My mother, Herminda, my brother, David, and my grandmother, Natividad, to whom this thesis is dedicated to, have been a constant source of love, support and strength that I have felt all this time even in the distance. Nada de esto hubiese sido posible sin el amor y paciencia de mi familia. Mi mamá, Herminda, mi hermano, David, y mi nona Tiva, a quienes dedico esta tesis, han sido una constante fuente de amor, sostén y fuerza durante todo este tiempo sin importar la distancia. Gracias querida familia por el apoyo espiritual durante todo este proceso y en mi vida en general. Su amor y confianza es el motor más importante en mi vida para seguir adelante, los amo inmensamente!

Most importantly, I'm deeply grateful to you Álvaro, mi amor, my best companion, my soul mate. Thank you for making me laugh everyday, for living every single minute of this journey with me, for being there through the thick and thin and for giving me the support, strength and motivation to do this. No me alcanzan las palabras para expresar lo que significas para mí, eres mi compañero en el viaje más importante de mi vida, ese viaje de nunca acabar. Te amo mucho guapo!

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Aim	2
1.3	Outline and main results	4
2	The effect of grade retention on secondary school performance: Evidence from a natural experiment	11
2.1	Introduction	12
2.2	Related Literature	15
2.3	Institutional Background	17
2.3.1	The Colombian Educational System	17
2.3.2	The Policy Change - the FRP reform	17
2.4	Empirical Strategy	18
2.5	Data Overview	22
2.5.1	Sources of Information and Sample Selection	22
2.5.2	Common Trends	23
2.5.3	Summary Statistics	27
2.6	Results	27
2.6.1	The Effect of the FRP Regime on Schools' Test Scores	27
2.6.2	On the Non-Linear Effects of Grade Retention	29
2.6.3	Effects along the Test Scores' Distribution	30
2.6.4	Testing the Common Trend Assumption	34
2.6.5	Additional Robustness Checks	38
2.7	Potential Mechanisms	39
2.8	Concluding Remarks	43
	Appendices	44
3	The effect of grade retention on secondary school dropout: Evidence from a natural experiment	51
3.1	Introduction	52
3.2	Related Literature	55
3.3	Background	57
3.3.1	The Colombian Educational System	57
3.3.2	The Free Retention Policy (FRP) reform	59

3.4	Empirical Strategy	59
3.5	Data and Descriptive Statistics	63
3.5.1	Data	63
3.5.2	Descriptive Statistics	66
3.5.3	Common Trends: Graphical evidence	68
3.6	Results	73
3.6.1	End-of-year dropout rates: Testing the common trend assumption and anticipatory effect	73
3.6.2	End-of-year dropout rates: The effect of the FRP reform and increased retention	75
3.6.3	Early dropout rates: Testing the common trend assumption and anticipatory effects	80
3.6.4	Early dropout rates: The effect of the FRP reform and increased retention	83
3.7	Conclusions	89
	Appendices	91
4	Does informal learning at work differ between temporary and permanent workers? Evidence from 20 OECD countries	105
4.1	Introduction	106
4.2	Related Literature	108
4.3	Model and Empirical Strategy	110
4.4	Data and Descriptive Statistics	113
4.4.1	Data and Sample	113
4.4.2	Variables	115
4.5	Empirical Results	122
4.5.1	On-the-Job Informal Learning Intensity	122
4.5.2	Robustness of Main Results	125
4.5.3	Informal Learning and Training: Substitution or Complementarity?	126
4.5.4	Heterogeneous Effects in Informal Learning	131
4.6	Conclusions and discussion	139
	Appendices	143
5	Work-related learning and skill development in Europe: Does initial skill mismatch matter?	149
5.1	Introduction	150
5.2	Related Literature	152
5.2.1	Human Capital Investments and Skill Development	152
5.2.2	Skill mismatch and Human Capital Investment	153
5.3	Data and Descriptive Analyses	156

5.3.1	Data and Sample	156
5.3.2	Variables and Descriptive Analyses	156
5.4	Estimation Method	162
5.5	Estimation Results	163
5.5.1	Work-Related Learning and Skill Development	163
5.5.2	On-the-job Learning Complementarity and Workers' Skill Development	176
5.6	Conclusions	182
	Appendices	184
6	Conclusion	195
	Valorisation Addendum	199
	Bibliography	205
	About the author	219
	ROA Dissertation Series	221

Introduction

1.1 Motivation

This thesis offers new insights into various factors related to the human capital development of children (at school) and adults (at work) over the lifecycle. The underlying tenet of this work is that human capital development contributes to both individual and collective economic and social success during initial education and henceforward. This thesis therefore emphasises the importance of formal education as well as school peers, families, co-workers and firms as sources of learning. This suggests that learning also occurs outside schools, especially on the job, which becomes an imperative source of skill development over the lifecycle as large as the investment in education (Heckman, Lochner, and Taber, 1998; Mincer, 1974).

Since the seminal work of Ben-Porath (1967), lifetime development of human capital is viewed as a process of skill formation, transmission, depreciation, restoration, replenishment, and growth. The studies contained in this thesis are all empirical contributions to the human capital literature, in which skill development is understood as a lifetime dynamic process with synergistic components that expand across both school and work. A key feature of this skill development process is that skills bolster each other over the lifecycle in a multiplier process (Cunha and Heckman, 2007).

Rapid technical change, globalisation and liberalisation have prompted policy makers in both developed and developing countries to prioritise skills development as a key strategy for economic and social progress. This because skills impact people's economic and social lives in many ways, not only in the labour market via labour productivity and career prospects but also also in a wide range of contexts such as organisational change, health, and social mobility, among others.

This thesis provides policy makers, schools' directors and human resource managers in firms with new insights into how to foster skill development. Because learning and skill development start in early infancy and continue throughout life, early learning begets later learning and early success may breed later success just as early failure may breed later failure. The general lesson from both the theoretical and empirical evidence is then that the later in life we attempt to

repair early skill deficits, or the more exclusive focus is put on formal education, the more costly skill development and remediation may become.

The first two studies in this thesis contribute to a better understanding of the positive and negative effects as well as the direct and spill-over effects of school failure on children’s human capital accumulation. The last two studies in this dissertation offer new insights into the determinants of human capital accumulation in the workplace – via both formal and informal learning – and the impact of such learning on workers skill development and depreciation. Once we understand and recognise the importance of learning failure and informal sources of learning for skill formation, we can think of policies to foster skill in different ways.

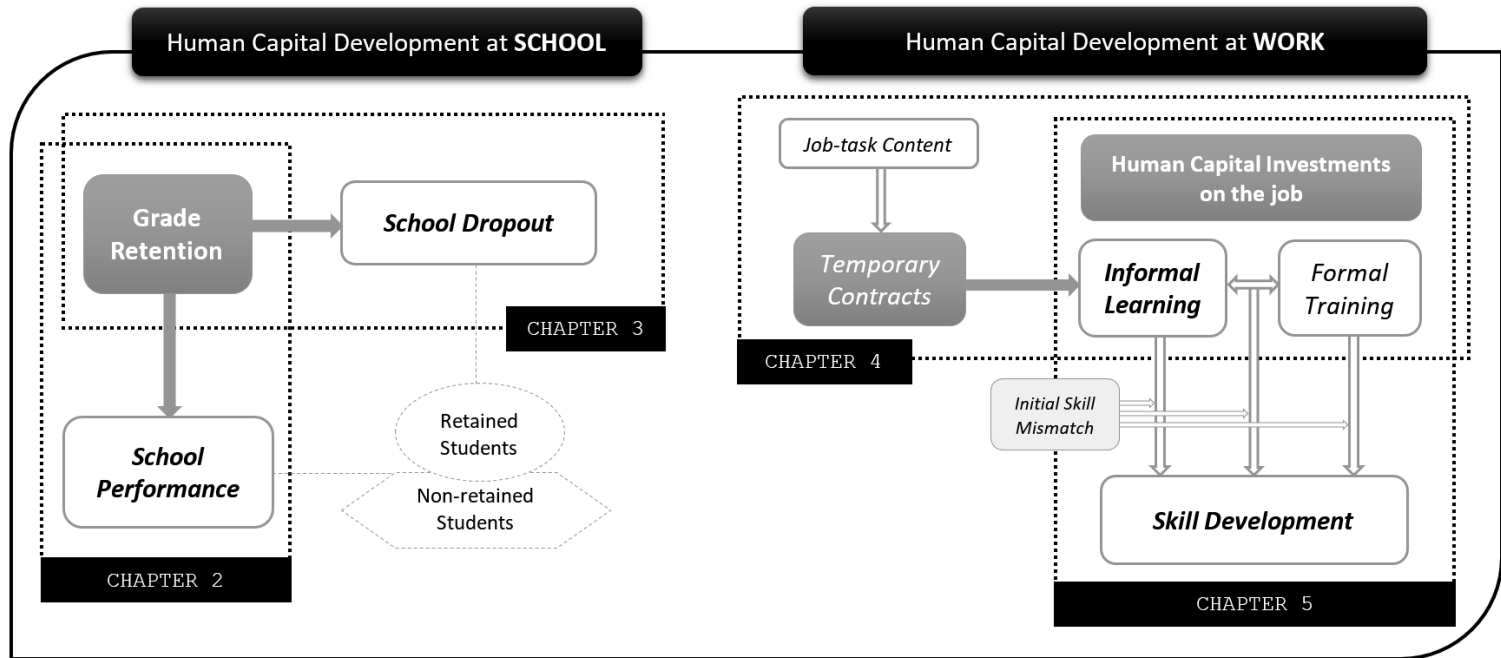
1.2 Aim

The aim of this thesis is to gain a better understanding of the dynamics and mechanisms of human capital development at school and work, looking at specific groups of individuals (i.e. students affected by grade retention and workers in temporary employment or in a job-skill mismatch). More specifically, this thesis contributes to filling some gaps in the human capital literature by addressing the following four related empirical research questions, articulated in the conceptual framework shown in Figure 1.1:

1. Does grade retention affect the school performance of retained and non-retained students?
2. Does grade retention influence school dropout of retained and non-retained students?
3. Does informal learning at work differ between temporary and permanent workers?
4. What is the relation between initial job-skill mismatch, work-related learning and skill development?

As Figure 1.1 indicates, the first two chapters of this thesis contribute to the literature on human capital development at school and the next two chapters to the literature on human capital development at work. The variables in black bold denote the four main outcomes of interest of this thesis: school performance; school dropout; informal learning on the job; and workers’ skill development. The variables in the grey boxes correspond to the main explanatory variables addressed in each of the chapters: children’s grade retention; temporary contracts; and human capital investments on the job. The filled grey arrows represent the relations we are able to plausibly identify in a causal framework whereas the white arrows represent standard estimated correlations.

Figure 1.1: Conceptual framework of the thesis



1.3 Outline and main results

This thesis consists of four self-contained chapters that deal with the four research questions presented in Section 1.2. The four chapters are all framed in the context of individuals' human capital accumulation and skill development over the lifecycle. Chapter 2 analyses the effects of grade retention on secondary school performance, and thereby tries to answer the questions whether children who repeat a grade improve their school performance, and whether grade repetition has any spill-over impact on the non-retained peer students. Chapter 3 investigates the effect of grade retention on school dropout by estimating the impact of grade failure on the incidence of dropouts among retained students at the end of the year of their retention as well as the effect on all students in the class the year after retention. Chapter 4 examines the influence of temporary employment on employees' informal learning at work to address the question whether workers with a temporary or permanent contract differ in the intensity of their human capital investments on the job. Finally, Chapter 5 provides new insights into the relation between work-related learning (formal training and informal learning) and employees' skill development on the job, considering the heterogeneity of this relationship with respect to workers' skill mismatch at job entry. Chapter 6 concludes.

Chapter 2: Grade retention and school performance

This study deals with the relationship between grade retention and academic performance of high-school students. The theoretical literature dealing with the effects of school retention and grade repetition on children's human capital accumulation is ambiguous regarding the net direct effects and the spill-over effects of retention. Moreover, the – mainly correlational – empirical literature shows mixed results, documenting both positive and negative estimates and mostly for retained pupils only.

The rationale followed in this chapter indicates that naïve estimations of the effect of retention on academic achievement may be negatively (positively) biased for retained (non-retained) students. To overcome the endogeneity problem, we exploit a policy change with respect to retention in Colombia. Until 2009, schools were each year restricted by national regulation to retain up to a maximum of 5 percent of their students. This restriction was ended by the Ministry of Education through a new mandate, allowing schools from 2010 onwards to retain as many students as they considered necessary. We use the term *Free Retention Policy* reform to refer to this policy change.

This study proposes a difference-in-difference framework to find the causal effect of grade retention on secondary school performance, distinguishing between retained and non-retained students. Specifically, it analyses the effect of students' retention in grade 10 on the math and language test scores of retained and non-retained students in grade 11, the last year of secondary school in Colombia. We exploit the variation in schools' retention rates induced by the policy reform and use two linked administrative datasets including students'

scores from the centralised high-school exit exam and school-grade retention rates. We find that retained (non-retained) students obtained higher (lower) language scores. In contrast, we observe no significant effects on math scores. We find that especially low-performing retained students benefited from increased retention. For non-retained students, the effects of increased retention are negative on language scores, especially at the lower end of the test performance distribution.

The main contribution of this study is the analysis of the non-linear effect of retention on test scores at various moments of the ability distribution. We show that modest increases in retention lead to higher scores in language for the retained students, but when many students are retained such gains decrease. These results imply that retention has marginal decreasing performance returns and, therefore, it is important for schools to retain students at optimal levels.

Finally, we provide evidence that neither average class size nor teachers' average educational achievements are the mechanisms that explain our main findings. In contrast, recruiting more high-quality new teachers in the school seems to be the channel through which increased retention affects school performance.

Chapter 3: Grade retention and school dropout

This chapter estimates the effect of retention on the dropout rates of secondary school students. The findings contribute to the literature that has consistently found that grade repetition is often associated to low enrolment, low graduation rates and high dropout rates (e.g. Bowers, Sprott, and Taff, 2012; Jimerson, Anderson, and Whipple, 2002; Roderick, 1994; Stearns, Moller, Blau, and Pottochnick, 2007). An important caveat in the literature is that the largest part of the evidence remains strictly correlational. In addition, the possibility that the strength of the association between retention and dropout may differ depending on the timing of retention during the school cycle has been largely overlooked.

This study contributes to filling the gaps in the literature in three ways. First, we propose a difference-in-differences framework to find the causal effect of retention on secondary school dropout rates, distinguishing between 1) the immediate effect on retained students who drop out of school by the end of the year of their retention, and 2) the dropout effect on all students enrolled in school the year after retention. Second, we analyse the heterogeneity of these effects at different moments of the education cycle by estimating the impact of retention on dropouts at each grade of secondary school from grade 6 to 11. Last, we are able to examine the heterogeneity and (non)linearity of the main effects with regard to treatment intensity by distinguishing between students in schools that reacted differently to the policy reform.

To analyse the effects of grade retention on early school leaving, we evaluate the dropout effects of the Colombian *Free Retention Policy* reform exploited in Chapter 2, using administrative school-grade records across nine school years and exploiting variation in retention rates at the school level, as induced by the afore mentioned reform.

Two major findings highlight from this chapter. First, a remarkably positive large effect of grade retention on end-of-year dropout rates among retained students and a positive, although relatively small, effect of grade failure on consecutive early dropout rates among all students enrolled in education. Second, we show significant heterogeneity in both effects depending on the timing of retention along the secondary education cycle: the effects of grade failure on early dropout rates are stronger when retention takes place at the earlier grades whereas the effect among retained students is strongest if retention occurs at grade 9 and grade 11, precisely the grades which successful completion entitles the students to receive the lower secondary school certificate and the high-school diploma respectively. This suggests there is a high cost of retention at these grades due to the forgone opportunity to finalise one of the two school levels of secondary education. We also show that both effects seem to be linearly proportional to the extent of retention growth, i.e., the strongest effect on dropout rates was experienced by students enrolled in schools where retention rates increased the most.

Chapter 4: Temporary contracts and informal learning at work

This chapter analyses the relationship between temporary employment and the intensity of on-the-job informal learning. The study builds on human capital theory that predicts that both firms and employees are less willing to invest in skills if workers are hired under temporary contracts. Although years of experience (or tenure) have played an important role in the economic literature as a proxy for unobservable investments in learning while working (Mincer, 1974), the empirical question on whether and, if so, to what extent learning informally on the job differs between temporary and permanent employees still remains.

With this study, we contribute to filling this gap by providing empirical evidence on the influence of temporary contracts on the intensity of informal learning at work across 20 OECD countries. For our analysis, we use unique data from the OECD Programme for International Assessment of Adult Competences (PIAAC) survey conducted in 2011 and 2012 and we estimate an instrumented endogenous switching regression model.

We contribute to the literature on flexible employment and human capital development in three ways. First, we estimate the extent to which the intensity of informal learning on the job differs between workers with temporary and permanent contracts. In doing so, we raise the issue of potential endogeneity of enrolment in a temporary job due to selection based on unobservable characteristics and exploit exogenous variation in the exposure of workers of different ages to employment protection legislation (EPL) and potential unemployment. Our main results show that workers in temporary jobs invest more intensively in informal learning than their counterparts in permanent contracts do, although the former are, in line with the empirical human capital literature, less likely, on average, to participate in formal training activities.

Second, we explore the interaction between training and informal learning

to analyse whether there is substitution or complementarity between these two types of learning for both temporary and permanent employees. We find evidence of a relation of complementarity, regardless of the type of contract, which suggests that the higher informal learning investments of temporary workers do not substitute for the lack of formal training. Third, we provide marginal treatment effects (MTEs) estimates to analyse the heterogeneity in workers' informal learning along the distribution of their individual unobserved characteristics. We find that the positive effect of a temporary placement on informal learning at work is expected to be larger among workers with lower propensities of selection into temporary jobs, that is, those who are likely to have better unobservable characteristics (e.g. ability and motivation).

Finally, this chapter provides additional insights on the possible mechanisms that could explain the main result. We discuss some heterogeneous-effect analyses that suggests that early career expectations of gaining a permanent contract could explain the higher informal learning investments of employees while in a temporary job. This explanation is informed by the theoretical idea that expectations of transition to more secure employment could lead to greater investments in human capital under uncertainty.

Chapter 5: Work-related learning, skill mismatch and skill development

This chapter builds on the seminal idea from human capital theory that lifelong learning improves workers' skills (Becker, 1964; Ben-Porath, 1967; Heckman, 1976; Mincer, 1962). Using data on more than 37,000 employees from the 2014 European Skills and Jobs Survey, this study contributes to the literature in four ways.

First, it provides more insights into the assumption that the productivity of job-related training is driven by the improvement of workers' skills by analysing the extent to which both training and informal learning on the job are related to employee skill development. Second, whereas most empirical studies focus on training participation, we are able to distinguish both formal training and informal learning. This enables us to investigate in more detail the extent to which workers' participation in different forms of work-related learning contributes in the improvement of their skill levels. Third, we provide more insights into the complementarity between training and informal learning on the job by analysing their interaction in the development of workers' skills. Fourth, we examine the heterogeneity of the relationships between training and informal learning and workers' skill development with respect to the workers' skill mismatch at job entry.

In line with the hypotheses derived from skills production models, we find that employees who participate in training or who are more often involved in informal learning show greater skill improvement. The relationship between informal learning on the job and employees' skill development appears to be stronger than that between training participation and skill development. However, there is complementarity between these two forms of work-related learning. This com-

plementarity seems to favour skill development since we find that workers who participated in both training and informal learning show significant additional improvement of their skills. This finding is consistent with the notions of the complementarity and cross-productivity of human capital (Cunha and Heckman, 2007) in an on-the-job context. We also find that skill mismatches induce heterogeneities and moderate the previous outcomes.

Thereby, we find that well-matched, underskilled, and overskilled employees differ in the extent to which investments in training and informal learning – and their complementarity – are associated with the accumulation of their skills during their working life. A stronger relationship between work-related learning and skill development suggests that initially underskilled employees benefit the most from both training and informal learning, whereas overskilled employees benefit the least. Human capital investments in the latter group seem to be more functional in counteracting skill depreciation and maintaining their skill level than in fostering skill accumulation, as suggested by literature on skill obsolescence (e.g. De Grip and Van Loo, 2002). Our study then suggests that being employed in a skill-challenging job or a job that underutilises a worker's skills has important implications on the returns to investment in both training and informal learning.

The effect of grade retention on secondary school performance: Evidence from a natural experiment

Abstract

In this study, we estimate the effects of grade retention on secondary school performance by considering a change in Colombia's educative legislation. In 2010, the rule that forced schools to retain no more than a 5 percent of their students was abolished. We exploit variation in schools' retention rates across time in a difference-in-differences framework and find that retained (non-retained) students improve (decline) their performance on language but not on math test scores. We suggest the school's position in the retention distribution, and the proportion of newly-hired teachers in the classroom, can be the mechanisms by which the marginally decreasing returns of grade retention are determined. We provide evidence that neither average class size nor teachers' average educational achievements are the mechanisms that explain our main findings. In contrast, recruiting more high-quality new teachers in the school seems to be the channel through which increased retention affects school performance.

JEL Classification: I20, I24, J24.

This chapter is joint work with Bart Golsteyn and Sergio Parra-Cely. This research is partly financed by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO). We are grateful to Lex Borghans, Giorgio Brunello, Andries de Grip, Caroline Hoxby, Annemarie Künn-Nelen, Steffen Künn, Olivier Marie, and participants of the Learning and Work Seminar Maastricht (2017), the IWAE Workshop on Applied Economics of Education (2017), and the conference of the European Society for Population Economics (2017) for their insightful comments and suggestions. We also thank the Colombian Inspectorate of Education (ICFES), the Colombian Bureau of Statistics (DANE), and the CEDE institute at the Universidad de los Andes (Colombia) for making available their datasets and school-identifiers on the SABER11 test and the C-600 census.

2.1 Introduction

Retention in school is common and widespread,¹ but its consequences for school performance are theoretically unclear and empirically diverse. Effects can be expected both for retained and non-retained students. For the retained, there may be positive effects as repeating a grade can help to acquire basic knowledge needed to perform well later on. But retention may instead also have negative effects on school performance if, for instance, self-esteem and motivation decrease as a result. For non-retained students, the relationship between grade retention and performance works via different mechanisms. In principle, students at the upper end of the ability distribution may learn more as the level of teaching increases if weaker students in class are retained. A positive effect of retention at the lower end of the ability distribution may be that the threat of being held back can stimulate children to work harder in school.² But this threat may also have negative consequences, as there is a negative relationship between mental stress and academic performance. Taken together, empirical research on the effect of retention on school performance is needed as its expected effects are ambiguous from a theoretical point of view.

An important empirical challenge in studying the relationship between grade retention and school performance is that omitted variables may drive the relationship. For instance, high ability children may be less likely to be retained and may also obtain higher school grades. This implies that a naïve estimation of the effect of retention on academic achievement may be negatively (positively) biased for retained (non-retained) students. In this chapter, we propose a framework to recover the causal effect of grade retention on secondary school performance of retained and non-retained students, which combines both administrative data and students' academic records. Specifically, we analyse the effect of retention in grade 10 on performance in grade 11, the last year in secondary school (nominal age: 16-17), using two administrative datasets from Colombia. The first, provided by the Inspectorate of Education, includes data for all students in the country on scores from a centralised exam in the last year of secondary education. The second dataset, from the Central Statistics Office, contains information on retention rates across all schools and grades in the education system. We are able to link the two datasets using unique school identifiers.

To overcome the endogeneity problem aforementioned, we exploit a policy change with respect to retention. From 2002 to 2009, under the automatic promotion policy regime, schools were by law not allowed to retain more than 5 percent of their students. In 2010, this directive was abolished and since then, schools are free to decide how many pupils should repeat a grade. The abolishment of the law increased retention rates dramatically in some schools, while

¹ In the United States, around 10 percent of all students are retained between kindergarten and eighth grade. In Germany and France, respectively 9 and 18 percent of all students are retained in primary school (Fruehwirth, Navarro, and Takahashi, 2016).

² One may also argue that this threat leads to other effects. Belot and Vandenberghe (2014) exploit a law reform to find that an enhanced threat of grade retention does not lead to better medium-term outcomes.

in others it had no effect. We use this information in a difference-in-differences approach, in which treatment and control groups are defined by the changes in retention attributed to the law change, analogue to the method used in recent papers in a different context than ours (Bauernschuster, Hener, and Rainer 2016; Havnes and Mogstad 2011). Schools in which retention rates increased more than the median change are labelled the treatment group and those that responded less than the median, the control group. Several placebo and falsification tests show that trends in math and language test scores are similar in control and treatment schools before the law was repealed, indicating that we can use a difference-in-differences model to estimate the causal effects of increased retention on school performance.

A key feature of our model is that we analyze the effects separately for retained and non-retained students. At the individual level, we do not have information on whether a student has been retained. However, we can identify the effects on retained and non-retained in the following way. We use information at the school level to define retained and non-retained students. An important piece of information here is that virtually no students are retained in grade 11. Recall that retention rates increased in treatment schools in 2010. This implies that one year later, in 2011, only the students who passed grade 10 will be taking the final exam in grade 11. Therefore, exam grades in 2011 are the grades of the non-retained students. One year later, in 2012, students who were retained for one year in grade 10 and students of the next cohort who were not retained in grade 10 will be taking the final exam in grade 11. Therefore, exam grades in 2012 are grades for a mixed group of non-retained and retained students. In our regressions, we analyse the effect of the new law on exam grades in both years simultaneously using lags. Controlling for the non-retained group (the first lag), allows us to identify the retained students within the mixed group (i.e., the second lag gives the effect on the retained and non-retained students minus the effect on the non-retained students). So we identify (1) the non-retained students with the first lag, and (2) the retained students with the second lag.

Using the definition of retained and non-retained students, as explained above, our findings indicate that higher retention in a school decreases language test scores for non-retained students, while it increases the scores of retained students. We observe no significant effects on math scores.

When analysing the results across the distribution of language scores we find that especially low performing, retained students benefited from increased retention. These results suggest that by repeating a class, students at the lower end of the ability distribution get a more thorough understanding of the material which enables them to perform better later on. Distinguishing between low, middle and highly treated schools reveals the non-linearity of the effect of retention: in middle treated schools, i.e. schools that moderately increased retention, the improvement is more pronounced than in highly treated schools. These results imply that retention has marginally decreasing performance returns and, therefore, it is important for schools to retain students at optimal levels.

For non-retained students, the effects of increased retention are negative on language scores, especially at the lower end of the test performance distribution. These results remain both qualitatively and quantitatively robust to the inclu-

sion of potentially important control variables, clustering of standard errors at the school level, and when performing other robustness checks. Theoretically, there are several reasons for this negative effect. Firstly, there might be a positive selection effect for non-retained students that is dominated by the plausibly negative influence of less able peers in the classroom. Secondly, students may strategically substitute effort between stem and non-stem subjects as the probability of repeating a grade rises. Decreasing marginal productivity in both courses implies that math scores are not expected to increase as much as language scores decrease.

While the above explanations are plausible, we cannot test them as our sources of information do not contain data on individual retention outcomes, intrinsic or extrinsic motivation, and/or individual effort across academic subjects. Instead, our analysis is focused on school-driven mechanisms that might explain the impacts we identify for retained and non-retained students alike. We provide evidence that neither average class size nor teachers' average educational achievements are relevant to explain these findings. In contrast, recruiting more high-quality new teachers in the school seems to be the amplifying force behind the benefits and costs of increased retention.

The main contribution of our study is that we analyse the non-linearity of the effect of retention on test scores at various moments of the ability distribution. We show that modest increases in retention lead to higher scores in language for the retained students, but when many students are retained, such gains decrease.

This analysis further contributes to the literature studying the effects of retention on educational outcomes. First, we analyse separately the effects of retention for retained and non-retained students. As indicated earlier on, the expected effects of retention are different for these groups. To evaluate the costs or benefits of retention for society, it is important to take the effects for both groups into consideration. Second, we also test transmission mechanisms at the school level, highlighting the role of teachers' staff composition on determining the differentials in test scores we observe as an outcome of increased retention. Third, our empirical approach to elicit causal effects departs from most other papers in this literature. We exploit the effects of a law change, which enabled schools to retain more children. Finally, we provide evidence on the effect of retention for a developing economy using a large administrative dataset, representative of the Colombian educational system.

The setup of this chapter is as follows. Section 2.2 discusses the literature related to our research question. Section 2.3 summarizes the Colombian context and the educational reform we exploit. Section 2.4 discusses the empirical strategy in detail. Section 2.5 describes our main sources of information, and the final dataset. Our central findings, including relevant robustness checks, are presented in Section 2.6. Section 2.7 offers a discussion on the transmission mechanisms. Finally, Section 2.8 concludes.

2.2 Related Literature

Our analysis contributes to the literature in various disciplines that have studied the effects of retention on school performance. Several articles in *School Psychology* and *Sociology of Education* analyse the relationship between grade retention and later school performance, mostly reporting this relationship to be negative. McCoy and Reynolds (1999) report that retention has a negative relationship with reading achievement. Jimerson, Carlson, Rotert, Egeland, and Sroufe (1997) find no evidence that retention is related to school performance. Jimerson (1999) follows students for 21 years in a longitudinal study to show that retained students have worse educational and employment outcomes in late adolescence. Silbergliitt, Appleton, Burns, and Jimerson (2006) find that retained students made less educational progress compared to a random group of other students. Stearns et al. (2007) report that students who repeat a grade prior to high school have a higher risk of dropping out of high school than students who are continuously promoted. An important caveat is that these articles report correlations and not causal estimates. Although correlations are informative, important confounders may bias such estimates. As previously explained, we expect a downward (upward) bias for retained (non-retained) students.

There is a small but growing literature that estimates the causal effect of grade retention on subsequent educational outcomes.³ The literature reveals that results are mixed, documenting positive as well as negative estimates. The results depend on the context and age of students. Firstly, some papers study the effects of retention at young ages. Koppensteiner (2014) examines the effect of automatic grade promotion on academic achievement (math scores) at primary school in Brazil. Applying a difference-in-differences approach that exploits variation over time and across schools in the grade promotion regime, the author finds a negative and significant effect of about seven percent of a standard deviation on math test scores. Fruehwirth et al. (2016) evaluate the effect of retention on achievement using data from children in kindergarten. Accounting for dynamic selection into retention, they find that children who are retained in kindergarten would have performed as much as 27 percent higher on math and reading tests in the next year if they had not been retained. Jacob and Lefgren (2004) instead find positive effects of retention at an early age. They assess the effects of retention in the Chicago Public School system using variation in retention generated by a test-based promotion policy, and find that retention has a modest but positive net impact on test scores for third grade students, while it increases academic achievement for low-achieving third graders. However, they also find that retention appears to have little or no effects for sixth-grade students.

Secondly, some studies have assessed the effects of retention on achievement

³ There may also be peer effects of retention. Hill (2014) investigates the extent to which course repeaters in high school mathematics courses exert negative externalities on their course-mates. Using individual and school-specific course fixed effects to control for ability and course selection, the study shows that increasing the share of repeaters in each course results in a moderate, significant increase in the probability of course failure for first-time course-takers. Results suggest that the negative effect is only evident when the share of repeaters reaches a threshold of 5 to 10 percent of the total number of course-takers.

in high school. A first set of papers reports negative effects. Jacob and Lefgren (2009) show that retention among younger students (sixth grade) does not affect the likelihood of high school completion, but retaining low-achieving eighth grade students in elementary school increases the probability that these students will drop out of high school. Manacorda (2012) studies the effects of retention in secondary junior high school (grade 7 to grade 9) in Uruguay on dropout rates and school attainment, exploiting a discontinuity established by a rule of automatic grade failure for pupils with more than three failed subjects at the end of the school year. The analysis reveals that retention increases school dropout and reduces school attainment. While analyses in secondary education focus mostly on dropout rates or completion of school, García-Pérez, Hidalgo-Hidalgo, and Robles-Zurita (2014) measure the effect of grade retention on Spanish students' PISA math scores at age 15, using the student's quarter of birth as an instrumental variable. They find that grade retention has a negative impact on educational outcomes. Those who are retained during primary education suffer more than those retained in secondary school.

Contrary to these findings, a second set of papers provides estimates of positive effects of retention in high school. Mahjoub (2017) finds large positive effects of retention on test scores, using quarter of birth as an instrument. The average effect of the treatment on the treated (ATT) ranges between one and one-quarter of a standard deviation of the test scores. Grade repetition in junior high school is also shown to increase the probability of graduation by 2.5 percentage points. Eide and Showalter (2001) use an instrumental variable for retention based on exogenous variation across states in kindergarten entry dates to find tentative evidence that retention may benefit students by both lowering dropout rates and raising labour market earnings. They find these effects to be relevant for white students, but not for black students.

A common approach in these studies is that the benefits of retention are evaluated at the margin where retention was increased by the natural experiment. An important issue with this approach is that the estimated benefits may differ at other moments of the ability distribution of students. For low performing students, the benefits of repeating a class may be positive while for high performing students there are probably negative effects. Schools are aware of this and aim to retain students until the marginal student does not benefit from retention.

Closest to our approach is the analysis developed by Koppensteiner (2014), who examines the effect of automatic grade promotion on academic achievement (math scores) in primary schools in Brazil using a difference-in-differences approach. Besides that we evaluate effects of retention separately for retained and non-retained students, and that we study the effects in secondary education and not in primary education, our study differs from his in the sense that we can show with placebo tests that the pre-treatment trends in school performance are common in treatment and control groups, i.e. the key underlying assumption of the difference-in-differences framework. Koppensteiner (2014) shows instead that school and student characteristics of treatment and control groups tend to be similar before the treatment occurred.

2.3 Institutional Background

2.3.1 The Colombian Educational System

Colombia has an eleven-year system of elementary and secondary education, consisting of five years of primary school (grade 1 to grade 5), four years of lower secondary education (grade 6 to grade 9) and two years of upper secondary education (grade 10 and grade 11).⁴ The expected age of entry to grade 1 is six years.⁵ Therefore, if children are not retained, they are expected to complete their secondary education at ages 16-17.

The educational system in Colombia is a comprehensive school system with no academic tracking at any grade.⁶ However, at the start of upper secondary education, schools differentiate in the provision of additional courses to complement the compulsory curriculum set by the Ministry of Education. These additional courses are organised in two specialisation programs: one is more academic and the other more technical in nature. The academic program provides general education in arts, sciences and humanities, whereas the technical program provides vocational knowledge and practice in technology, craft industry, business, pedagogics, or agriculture.

Upon completion of the grade 11 of secondary school, all students, regardless of the chosen program, participate in a national standardised exam (“SABER11,” in Spanish), an achievement and competency test that is administered every year by the National Institute for the Assessment of Education (“ICFES,” in Spanish).⁷ This exam is a high-stakes evaluation, required not only for admission to tertiary education, but also to receive the high-school diploma. This test is also widely considered as the reference examination to evaluate the quality of secondary education across the country. In line with previous literature on grade retention, we focus on students’ performance on the math and language parts of the test as the main outcome of our analyses.

2.3.2 The Policy Change - the FRP reform

In 2002, by mandate of the Ministry of Education (Decree 230 of 2002), schools were each year permitted to retain up to a maximum of five percent of their students. This retention policy was implemented to reduce costs attributed to higher retention rates (i.e. low performance, low motivation, etc.) without compromising the quality of education provided by the system (Martínez

⁴ Elementary and secondary education in Colombia is offered in two school calendars: A calendar labelled “A” that runs from February until November, and a calendar “B” from September to June. Most schools (92%) in the country operate in calendar A. Formal education is also offered by schools in three different class-schedules: a morning schedule, an afternoon schedule, and a full-day schedule. Students opt or are allocated by the school to attend either one of these. Most students in secondary education attend school either in the morning or the afternoon schedule (78%).

⁵ This age is suggested but not mandatory since in Colombia there are no compulsory age-at-school entry laws.

⁶ The Ministry of Education regulates all levels of education and national exams for both publicly and privately funded schools.

⁷ Hereafter, we will refer to this institute as the Inspectorate of Education.

and Herrera, 2002). According to the policy mandate, a student should have been retained if at least one of the following three circumstances occurred: i) the student received an unsatisfactory performance evaluation in three or more academic subjects in the current academic year, ii) the student received an unsatisfactory performance evaluation in math and/or language courses during the current and two previous grades, or iii) the student failed to attend at least 25% of all academic activities during the current academic year. However, schools were required to adjust their evaluation standards to comply with the law, which forced them to promote at least 95% of all their students.

While the reform was considered as moderately successful in terms of reducing school dropouts, the incentives to underperform at school as perceived by schools, teachers, and parents, led the Ministry of Education to revoke the Automatic Promotion Policy Rule.⁸ In February 2009, the 5-percent retention restriction was replaced by the Ministry of Education through a new regulation mandate (Decree 1290 of 2009), allowing schools from 2010 onwards to retain as many students as they considered necessary, and thereby giving them more discretion in their evaluation and promotion procedures. We use the terms Automatic Promotion Policy (AUP) to indicate the period until 2009 and Free Retention Policy (FRP) from 2010 onwards. Overall, the abolishment of the AUP regime increased students' retention rates across all grades of secondary education from 4.3 percent to 7.7 percent, on average, in all schools in the country.

2.4 Empirical Strategy

We evaluate the effects of retention in grade 10 on math and language performance in the secondary school exit exam conducted at grade 11. The empirical challenge in studying this question is that omitted variables may drive the relationship. A naïve estimation using OLS may be negatively biased for retained students if the lower scores they obtain are not due to retention but to their lower ability. As students' ability increases, we might expect the benefits of grade retention to be decreasing and, for the upper end of the ability distribution, to negatively impact academic performance. Nonetheless, such counterproductive effects may be veiled by, for instance, the positive sorting of skilled students in subsequent grades as a byproduct of increased retention.

We exploit the before-mentioned policy change in Colombia that occurred in 2010 and implied that schools facing constraints in their retention requirements (AUP regime) were now allowed to retain as many students as they considered appropriate (FRP regime). To identify the effect of grade retention on test scores, we implement a difference-in-differences framework which exploits the school-year variation in retention rates.

Schools reacted differently to the new policy, suggesting that the grade retention effect is heterogeneous across schools. We classify schools in two groups: the treated group, consisting of schools that increased their retention rates after the law change, and the control group, composed by schools in which retention

⁸ Ministry of Education, Press Release April 17, 2009.

rates remained relatively constant. We classify schools into the treated or control categories using the difference between the schools' average retention rate at grade 10 between both policy regimes, the AUP regime (2007-2009) and the FRP regime (2010-2012). Sorting schools on this difference, we define the treated group as the pool of schools with an above-median increase in their retention rates, and the remaining schools are labelled the control group.⁹ Panel (a) in Figure 2.1 shows the retention rates for treated and control schools across years. In control schools, on average, these rates increased slightly after the law change by approximately 0.4 percentage points. In contrast, treated schools increased retention rates by 7.6 percentage points, implying that the latter retained 7.2 percentage points more than control schools.

Additionally, we classify treated schools based on quintiles of the difference in retention rates. Panel (b) in Figure 2.1 shows the retention rates across time for four quintile groups.¹⁰ In which schools raise retention during the FRP regime. First, we observe that retention rates among schools in the second quintile barely change. These schools can be considered as an alternative control group. Second, we observe three groups of schools (quintiles 3 to 5) that are affected differently by the policy change. Since retention rates in these groups increase on average by 2, 6, and 12 percentage points, we label these schools as low treated, medium treated, and highly treated, respectively. With the exception of this latter group, all remaining schools retained students in the AUP regime as required, with an average retention rate of 3.6 percent.

The baseline difference-in-differences specification we implement is:

$$Y_{st} = \alpha_s + \delta_t + \sum_{h=1}^2 \gamma_h [Group_s \times FRP_{t-h}] + \beta X_{st} + \varepsilon_{st}, \quad (2.1)$$

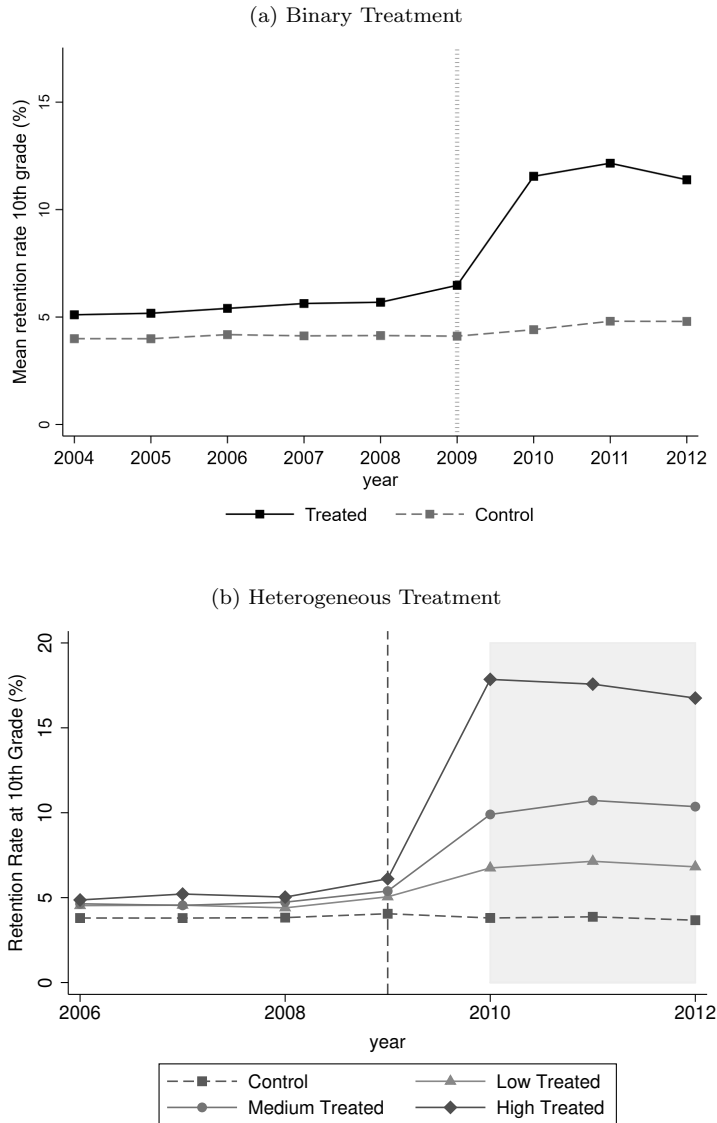
in which Y_{st} denotes standardised test scores for school s in exam year t . α_s and δ_t are fixed effects by school and exam year, respectively. $Group_s$ is a dummy variable that takes value 1 for schools in the treated group, and zero for schools in the control group.

In our basic specification, treated schools ($Treated_s$) are those with above median changes in retention rates and control schools are those with a below-median change in retention rates. In our more elaborate specification, we include three treatment dummies corresponding to low treated ($LowTreated_s$), medium treated ($MiddleTreated_s$), and high treated schools ($HighTreated_s$). FRP_{t-h} is an indicator variable with value 1 if the FRP regime was in place h years

⁹ In the economics literature, this treatment-control classification is also implemented for instance by Havnes and Mogstad (2011) and Bauernschuster et al. (2016). These authors analyse the effects of increased child care coverage on parental economic outcomes.

¹⁰ The first quintile is excluded from this graph as this is a group of defiant schools four quintile groups in which schools raise retention during the FRP regime. which *decreased* retention rates by 4 percentage points. As we consider these schools not being fully comparable with the universe of compliant schools, we decided to exclude them only from this specific analysis. Hence, we keep all observations in our estimations, but we refrain from interpreting effects for schools in the defiant group. This restriction in our analyses is only relevant when we account for the heterogeneous effects of retention. In contrast, results involving treated and control schools classified by the use of the above-median increase in retention correspond to all schools in our sample.

Figure 2.1: Retention rates treatment and control groups



Notes: Control vs. Treated Schools. Panel (a) displays retention rates at grade 10 (in percentage points) for treated and control schools. Treatment schools are defined as those with above -median increase in their retention rates at grade 10 from the automatic promotion years to the free retention years. Control schools are defined as those with below -median increase. Control vs Treated Schools - Heterogeneous Treatment Status. Panel (b) shows retention rates at grade 10 for control and treated schools, for the different treatment definitions explained in the main text: Highly Treated ($HighTreated_s$), Medium Treated ($MiddleTreated_s$), and Low Treated ($LowTreated_s$). The dashed vertical line denotes year 2009 where schools were notified that the AUP regime will no longer hold. The gray area denotes the years where the FRP regime was in place.

before exam year t , and zero otherwise. The interaction term $Group_s \times FRP_{t-h}$ therefore measures the variation in tests scores that can be causally attributed to the shift in the retention policy from year $t - h$ onwards. For each regression, we run two specifications. First we run our regressions without covariates. Second, for each exam year we include covariates for the first two lags of school-specific attributes that change over time. In this way, we account for pre-FRP variation in characteristics among schools. This set of control variables is denoted in the equation as X_{st} . Finally, ε are standard errors clustered at the school level.

We aim to analyse the effects of increased retention separately for retained and non-retained students. We do not observe individual retention outcomes but we are able to recover school-level retention rates from administrative sources. As we use retention rates at grade 10, the pool of test-takers in our sample consists of non-retained students and students who were retained in grade 10 only once. Controlling for the first two lags of our “FRP regime exposure” variable, we can differentiate the effect for each type of pupil.

The cohort of students taking the exam in year t is largely composed of two types of students: i) students in grade 10 in year $t - 1$ that were promoted to grade 11 at year t , and ii) students in grade 10 that were retained in year $t - 2$, repeated and passed grade 10 in year $t - 1$, and enrolled to grade 11 in year t . Hence, our parameters of interest are γ_1 and γ_2 .

The first parameter measures the effect of being exposed one year to the FRP policy in grade 10 on schools’ tests scores the next year in grade 11. The expected direction of this effect is ambiguous. On the one hand, the sign of γ_1 reveals whether non-retained students benefited from higher retention rates because of a positive sorting effect. In this case, we expect the effect to be positive. On the other hand, if we interpret this coefficient as the effect of increased retention for the marginal student (i.e. students that should have been retained but were promoted at the margin), we might expect the impact to have the opposite sign relative to the effect of retention on the retained students, For instance, if the latter effect is positive, marginal students are worse off when promoted to grade 11 because they will miss the chance to receive further training on the academic subjects they struggled with the most. The second parameter γ_2 measures the impact of FRP regime’s exposure in the previous two consecutive years on schools’ test scores. Assuming that students are retained in grade 10 only once, this impact can be mostly attributed to retained students (the second lag gives the effect on the retained and non-retained students minus the effect on the non-retained students).

Because of our treatment-control classification, γ_1 and γ_2 are best interpreted as intention-to-treat effects (ITT). To obtain the average treatment effects on the treated (ATT), we will rescale these coefficients by the difference in retention rates between treated and control schools implied by the law change.

The main identification assumption in this setting is that the variation in retention rates is orthogonal to expected changes in test scores. This assumption is equivalent to claim that treatment and control schools would have shared similar trends in test scores if the retention policy had remained the same. We

formulate an alternative specification to test this assumption:

$$\begin{aligned}
Y_{st} = & \alpha_s + \delta_t + \sum_{k=2008}^{2010} \mu_k [Group_s \times (Year = k)] \\
& + \sum_{k=2011}^{2013} \theta_k [Group_s \times (Year = k)] + \beta X_{st} + \varepsilon_{st}. \quad (2.2)
\end{aligned}$$

In equation (2.2), the null hypothesis of interest is that pre-FRP differences in trends between treated and control units are not significantly different from zero (i.e. $\mu_{2008} = \mu_{2009} = \mu_{2010} = 0$). Namely, we control for the interaction between the treatment status and those exam years where test takers, by construction, were not exposed to increased retention rates because of the policy change. As we will elaborate further on, we are not able to reject such hypothesis at conventional significance levels.

In addition to the above specification, we also perform placebo tests to account for artificial policy changes that should not have any effect on test scores:

$$Y_{st} = \alpha_s + \delta_t + \sum_{h=1}^2 \pi_h [Group_s \times FakeFRP_{k,t-h}] + \beta X_{st} + \varepsilon_{st}. \quad (2.3)$$

In equation (2.3), $FakeFRP_{k,t-h}$ is an indicator variable that takes the value 1 if the FRP regime was in place during year $t-h$, assuming it (artificially) started either in $k = 2008$ or $k = 2009$. By not being able to reject the null hypothesis of non-significant effects (i.e. $\pi_1 = \pi_2 = 0$), we are confirming that the changes in test scores can be attributed to the elimination of the AUP regime only.

2.5 Data Overview

2.5.1 Sources of Information and Sample Selection

The sample we use in this chapter comes from two main sources. The first is an administrative dataset from the Colombian Inspectorate of Education. The Inspectorate provides freely downloadable micro-level data on the centralised exam conducted among 2.7 million pupils in their last year of secondary education (grade 11).¹¹ This exam, known as SABER11, is a standardised test that evaluates every year a range of seven school subjects.¹² Test scores range from 0 to 100 in each subject and they are standardised by subject at the national level, so that each student's score is informative about his/her position relative to the national average in that subject. According to the Inspectorate of Education, the tests are comparable for the period 2000-2013.

We use available data from 2007 to 2013 that include math and language scores; student and school identifiers; some schools' attributes such as the aca-

¹¹ This exam takes place every year in the month of September, three months prior to the official end of the school calendar A.

¹² These subjects are: Math, Language, Physics, Biology, Chemistry, History, and Philosophy.

demographic calendar, daily class schedule, public or private status, specialisation programmes offered; and information on several individual characteristics such as age, gender, mother’s education, and other socio-demographic indicators. We collapse these data at the school-year level, and focus only on our outcomes of interest (i.e. math and language test scores) at several moments of the distribution: mean, 10th percentile, 25th percentile, median, 75th percentile and 90th percentile of the schools’ test scores. These scores are all standardised over the entire sample to interpret the effects in terms of standard deviations (SD) at the school level. While representative of the student body that is assessed at the last year of secondary school, this dataset does not collect information on pupil’s retention at any stage of the education process.

To obtain retention rates at the school level we rely on the schools’ official census, which the Ministry of Education releases each year for public use through the National Bureau of Statistics (DANE, in Spanish). Known as the C-600 census, this dataset contains information on academic indicators that all schools in the country are compelled to report on a yearly basis. We use information from this dataset on retention rates at grade 10, as well as other school characteristics, such as the number of groups per grade, the number of students enrolled at grade 10, the number of teachers with a professional degree, the number of teachers hired under the old and new pay scales regulated by the central government, and the number of non-academic staff (managerial, support, health) per school. We use this universal census information for the period 2005-2012. Using unique school identifiers, we are able to match 88.2% of all schools’ test scores to the respective school retention data for the entire period 2007-2013. These matched data correspond to 85.6% (N=2,363,997) of all students that took the exam during the same period.

Our unit of observation is a school-exam year combination. The estimation sample consists of first-time SABER11 test takers¹³ from schools that i) offered education exclusively in Calendar A (February to November), ii) did not change this calendar during the period 2007-2013, iii) had no missing values on tests scores, retention rates, and schools’ covariates, and iv) reported information on retention rates for at least 3 years, with at least one year before and after the retention law changed. The resulting dataset consist of an unbalanced panel of 6,248 schools, which in total across the 2007-2013 period contains 35,693 observations.

2.5.2 Common Trends

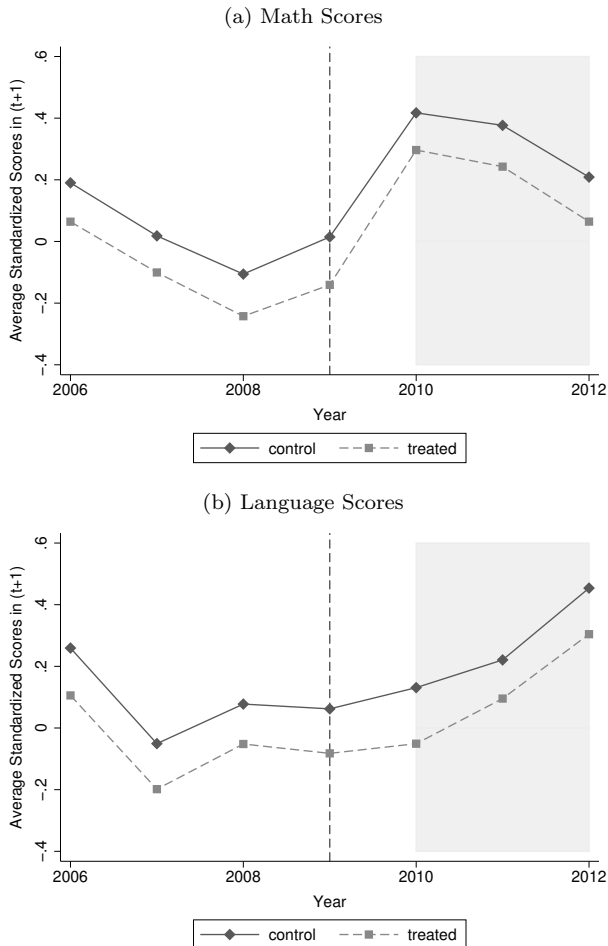
Figure 2.2 in panels (a) and (b) shows the average test scores across time for treated and control schools in math and language, respectively. On average, control schools performed better at both subjects during the AUP regime. For instance, students in control schools scored 0.1 of a standard deviation more in the math exam than students enrolled in treated schools. The same patterns are observed in the language exam with students scoring around 0.12 of a standard

¹³ We leave out of the sample the top 1% and bottom 1% of students in terms of their age reported at the exam. This selection criterion excludes extreme outliers who reported ages below 12 or above 40.

deviation higher than students in treated schools. Appendix Figure 2.A1 shows the same results when we plot the residuals from a regression including exam-year and school fixed effects, as well as time-variant school attributes. Figure 2.3 reveals that such trends are also common when distinguishing the treatment groups for high, middle, and low treated schools. These results remain robust to the inclusion of school-specific covariates (Appendix Figure 2.A2).

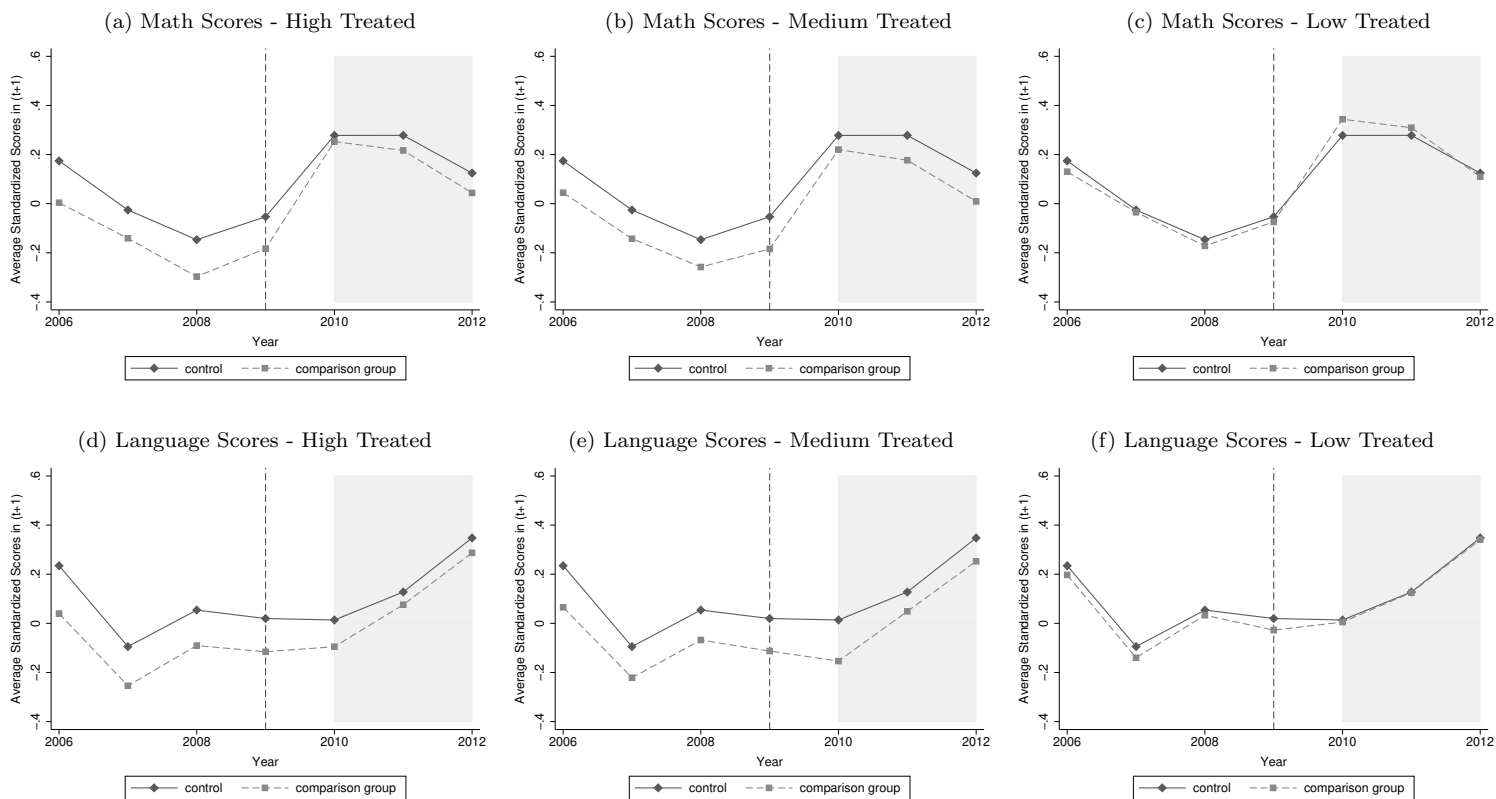
The main conclusion from these graphs is that there is a common trend in test scores between treated and control schools. This allows us to use a difference-in-differences strategy. The difference-in-differences estimator will isolate time invariant confounding factors, leaving the remaining variation to be attributed to the effect of increasing retention in schools. In the results section, we will provide robust statistical evidence that the common trend assumption holds.

Figure 2.2: Test Scores by Treatment Status



Notes: This figure displays average tests scores in year $(t + 1)$, for AUP and FRP years. Panel (a) presents common trends on average math scores between treated (dashed lines) and control (solid lines) schools. Panel (b) shows similar trends on average language test scores.

Figure 2.3: Test Scores by Treatment Status - Multiple Treatment Groups



Notes: This figure displays average test scores at year $(t + 1)$ in the AUP and FRP years for the multiple treatment groups described in the text. Panels (a)-(c) show average math scores. Panels (d)-(f) present the same figures for average language scores. The dashed vertical line denotes year 2009 where schools were informed that the AUP regime will no longer hold. The gray area denotes the years where the FRP regime was in place.

Table 2.1: Summary Statistics: Schools' Characteristics during AUP Regime

	Treated		Control		Both		
	# Schools (1)	Mean (2)	# Schools (3)	Mean (4)	Difference (2)-(4)	s.e.	
Average characteristics during AUP years (2007-2009)							
<u>Socio-Demographic School's Composition</u>							
Proportion of female students	3,142	0.528	3,106	0.527	0.001	0.004	
Proportion of students from rural areas	3,142	0.342	3,106	0.328	0.014	0.010	
Proportion of students from ethnic minorities	3,142	0.078	3,106	0.080	-0.002	0.005	
Average age at exam date	3,142	17.628	3,106	17.598	0.030*	0.016	
Proportion of students with an educated mother	3,142	0.149	3,106	0.224	-0.075***	0.006	
Proportion of students above poverty classification	3,142	0.253	3,106	0.343	-0.090***	0.008	
Public school	3,142	0.272	3,106	0.388	-0.117***	0.012	
Class schedule morning: 7:00 to 12:00	3,142	0.558	3,106	0.502	0.056***	0.012	
Class schedule afternoon: 13:00 to 18:00	3,142	0.170	3,106	0.110	0.061***	0.008	
School type: academic and technical	3,142	0.625	3,106	0.644	-0.019	0.012	
School type: pedagogical training	3,142	0.154	3,106	0.128	0.026***	0.009	
School type: technical	3,142	0.020	3,106	0.014	0.006*	0.003	
<u>School-Related Attributes</u>							
Average class size	3,142	0.201	3,106	0.214	-0.013	0.010	
Average number of groups at grade 10	3,142	32.272	3,106	29.491	2.781***	0.302	
# of teachers with qualifications	3,142	2.367	3,106	1.839	0.528***	0.047	
Proportion of teachers under the new pay scale	3,142	0.913	3,106	0.881	0.032***	0.003	
Average number of managerial personnel	3,142	0.159	3,106	0.129	0.030***	0.005	
Average number of support staff	3,142	3.204	3,106	2.837	0.367***	0.049	
Average number of health personnel	3,142	0.768	3,106	0.806	-0.038	0.029	
		3,142	0.166	3,106	0.239	-0.072***	0.015
Total Schools	6248						

Notes: Data on socio-demographic composition of schools comes from the ICFES SABER11 dataset. Data on schools' attributes come from the administrative records of the C600 school made by the national statistics office (DANE). Treated and control schools are defined as in the text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.3 Summary Statistics

Table 2.1 reports summary statistics for the sample used to estimate our main results. We present information on schools' characteristics during the AUP regime. Columns (1) and (3) report the number of schools per treatment status, and columns (2) and (4) present the averages of each control variable for both treated and control groups, respectively. Columns (5) and (6) report differences in means and standard errors between treated and control schools.

Schools differ systematically in their attributes during the AUP regime. Considering socio-demographic attributes of students, control schools present a more favorable composition of students from highly educated households (measured as the proportion of mothers with tertiary education), and few students with poverty status. There is also a larger proportion of public schools in this group, relative to the treatment group. However, treated schools operate under shorter working spells relative to schools in the control group. Moreover, treated schools also seem to present some academic differentiation as they also provide other types of training (e.g. pedagogical, technical vocational training).

With regards to school-related characteristics, treated schools, on average, have more groups per grade and more qualified teachers employed at school. Regarding teachers' compensation and renewal of personnel, we observe that treated schools hire slightly more staff under the new pay scale than control schools, but the overall proportion of teachers under the new pay scale increased during the last three years of the AUP regime for all schools. In contrast, control schools seem to employ more health professionals (e.g. dentists, physicians) than treated schools. Conversely, treated schools seem to hire more staff for managerial purposes than control schools.

While time-invariant differences between treated and control schools are controlled for by the inclusion of school-specific fixed effects, a potential concern for the identification strategy implemented in this chapter is that time-varying schools attributes change at the same time of the policy intervention. To address this issue, in some specifications, we include school-specific, time-varying attributes one and two years prior the exam date. As we will elaborate further in our results section, our estimates remain invariant to the inclusion of these controls, suggesting that we identify the effect of retention, net of other elements affecting test scores across time.

2.6 Results

2.6.1 The Effect of the FRP Regime on Schools' Test Scores

Table 2.2 presents our baseline estimates on the effects of higher retention due to the law change on average math and language standardised test scores. As implied by equation (2.1), in some specifications, we include a set of time variant school-specific attributes to obtain the net impact of the FRP regime. All standard errors are clustered at the school level to ensure we account for potential serial correlation, as indicated in the difference-in-differences literature (Bertrand, Duflo, and Mullainathan, 2004).

Our findings show that, across all years, the increase in retention does not have a meaningful economic effect on average math school performance. In contrast, we obtain a positive and strongly significant effect on language scores of 6.5% of a SD for a consecutive 2-year prior exposure to the FRP regime, and an average negative effect of 5.5% of a SD (Columns (4)-(5), first row) for being exposed to higher retention rates one year before the exam is taken. As treated schools increased retention 7.2 percentage points more than control schools, the effect implies an increase in language scores of 8.7% of a SD for a 10 percentage point rise in retention rates at grade 10.

Table 2.2: Effect of the FRP Regime on Average Test Scores

	Math Scores		Language Scores	
	(1)	(2)	(4)	(5)
$Treated_s \times FRP_{t-1}$	0.003 (0.020)	0.015 (0.020)	-0.061** (0.027)	-0.051* (0.027)
$Treated_s \times FRP_{t-2}$	0.002 (0.018)	0.006 (0.018)	0.064*** (0.023)	0.066*** (0.023)
Observations	35,693	35,693	35,693	35,693
R-squared	0.149	0.155	0.094	0.101
# Schools	6,248	6,248	6,248	6,248
Covariates	No	Yes	No	Yes

Notes: Robust standard errors clustered at the school level reported in parentheses. All specifications include fixed effects by school and exam year. Treated (Control) schools are defined as those with above median (below median) increase in their retention rates at grade 10 from the AUP years to the FRP years. The outcome variables are average standardised SABER11 test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. Covariates considered in these estimations include the first two lags of: Average class size at grade 10, number of health, support, and managerial (non-academic) staff per school, number of teachers with a professional degree, number of teachers under the new and old government-regulated pay scales, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

A plausible explanation for the positive effect attributed to retained students is that repeating a grade allows them to get a more thorough understanding of the material. Conversely, there may be several reasons for the negative effect on the non-retained students. First, the positive selection effect may be dominated by the negative influence of being in a group with a large fraction of lower performing peers after retention rates increased. Secondly, it may be that students started to allocate strategically more time and effort to study stem subjects when retention rates increased. Decreasing marginal productivity in both subjects implies that math scores do not increase as much as language scores decrease. This latter effect might be particularly relevant for students at the margin of repeating a grade.

2.6.2 On the Non-Linear Effects of Grade Retention

Table 2.3 documents the effects using the baseline specification implied by equation (2.1) with three dummy variables for all treatment groups of interest, i.e., low, middle and highly treated schools. In line with our basic specification, results from columns (1)-(2) suggest no significant effects of being exposed to the FRP period in years $t - 1$ and $t - 2$ on average math scores at year t . In contrast, we obtain significant effects for average language test scores which vary depending on the school's treatment classification (columns (3)-(4)).

Table 2.3: Effect of the FRP Regime on Average Test Scores by Multiple Treatment Status

	Math Scores		Language Scores	
	(1)	(2)	(4)	(5)
$HighTreated_s \times FRP_{t-1}$	0.037 (0.033)	0.053 (0.033)	-0.093** (0.043)	-0.081* (0.042)
$HighTreated_s \times FRP_{t-2}$	-0.007 (0.030)	-0.003 (0.031)	0.124*** (0.037)	0.127*** (0.037)
$MiddleTreated_s \times FRP_{t-1}$	-0.015 (0.032)	-0.003 (0.032)	-0.138*** (0.043)	-0.127*** (0.043)
$MiddleTreated_s \times FRP_{t-2}$	-0.017 (0.030)	-0.013 (0.030)	0.127*** (0.037)	0.128*** (0.037)
$LowTreated_s \times FRP_{t-1}$	0.020 (0.033)	0.027 (0.032)	-0.077* (0.046)	-0.072 (0.045)
$LowTreated_s \times FRP_{t-2}$	-0.014 (0.030)	-0.012 (0.031)	0.061 (0.039)	0.062 (0.039)
Observations	35,693	35,693	35,693	35,693
R-squared	0.149	0.156	0.095	0.102
# Schools	6,248	6,248	6,248	6,248
Covariates	No	Yes	No	Yes

Notes: Robust standard errors clustered at the school level reported in parentheses. Treated schools are defined as in the main text. The outcome variables are average standardised SABER11 test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. Covariates considered in these estimations include the first two lags of: Average class size at grade 10, number of health, support, and managerial non-academic staff per school, number of teachers with a professional degree, number of teachers under the new and old government-regulated pay scales, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Coefficients displayed in the first, third, and fifth rows suggest that middle treated schools score 13% of a SD lower relative to control schools. The effects for all other treatment groups, even those not statistically significant, suggest a non-linear parabolic pattern.

The increase in language scores of the retained students is significant both for the middle and highly treated groups. Interestingly, the gains on test scores appear to be similar for these groups. Dividing these results by the percentage

points increase in retention rates corresponding to each group implies that the ATT coefficient of the highly treated group is smaller than that of the medium treated group. Considering that highly treated schools experienced a 12 percentage point increase in retention because of the FRP regime, our findings imply that a one percentage point rise in retention rates at grade 10 explains a 1.08% of a SD increase in test scores. The aforementioned impacts are larger for schools in the “medium treated” group. Namely, a 6 percentage points increase in retention rates at year $t - 2$ implies a rise in test scores of about 13% of a SD in year t , so the effect is around 2.16% of a SD per one percentage point increase in retention. Hence, the same percentage point change in middle treated schools is twice as effective as it is in highly treated schools.

The non-linear effect on language scores suggests that retaining students is a strategy that exhibits decreasing marginal returns. At some point, higher retention is not expected to increase language performance. These students may, for instance, become demotivated because they must take the same classes again. For the marginal students, the results are only highly significant for the middle treated group which also shows the non-linear nature of the effects obtained.

The main conclusion from these findings is that there is a non-linear effect of retention, i.e. schools that retain many students do not necessarily benefit more from such a strategy, relative to other schools promoting more students. For the same percentage point increase in retention, middle treated schools obtain larger gains relative to schools that retain more students.

2.6.3 Effects along the Test Scores’ Distribution

In Figure 2.4 we plot our difference-in-differences coefficients, this time considering test scores’ percentiles by school as dependent variables, and using the above-median change in retention as the treatment classification criteria. In all these estimations, we include school-specific covariates, although results barely change when the latter are excluded.

For math scores, we again obtain no effects for retained and marginal students across the entire test score distribution. On the contrary, results on language scores attributed to marginal pupils displayed in panel (c) show that the negative effect discussed earlier is strongest at the lower end of the distribution. For example, students performing at the 25th percentile in treated schools score up to 7% of a SD lower than comparable students at control schools. As average language scores for all students in the upper segment of the distribution were not affected by retention, these findings suggest that the negative effect of being exposed to higher retention is more severe for lower performing non-retained students. We conjecture that this effect can be explained by the fact that students at the bottom of the ability distribution faced a higher threat of retention under the FRP regime, therefore compromising their test scores in the future.

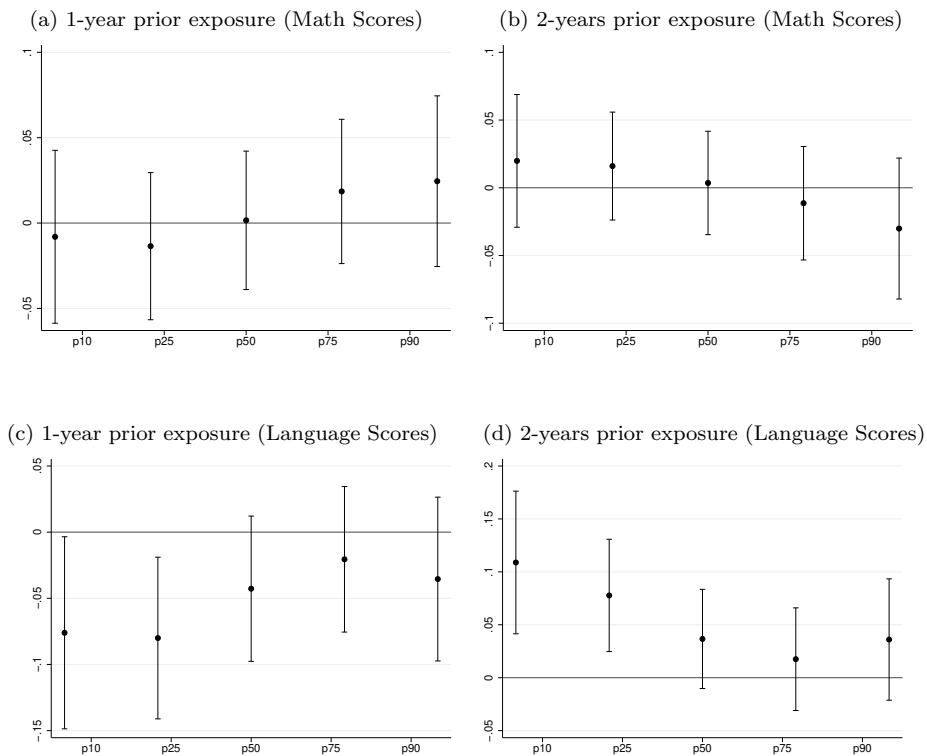
Regarding language scores of retained students, the effect of higher retention is decreasing in students’ test performance. Relative to students at the 10th percentile, students in treated schools scored 11% of a SD higher because of the FRP Regime. The effect is significant but reduces in magnitude as perfor-

mance increases. For instance, students at the 25th percentile score 8% of a SD more than students with comparable performance at control schools. The FRP regime appears to have no effect for those students performing at the median or above. Overall, these findings suggest that pupils at the lower end of the distribution benefited more from being in schools with increased retention rates due to repeating their coursework. This may also indicate that the benefits of retention are not linear. That is, being retained can be highly beneficial for underperforming students.

We perform the above exercise also for the specification in which we classify three treatment groups as already discussed. Table 2.4 presents the corresponding difference-in-differences coefficients for math and language scores. The table shows the results across the test scores' distributions distinguishing between low, medium and high treated schools. For math scores we obtain positive impacts at the upper end of the distribution that can be attributed to marginal students, although these effects are only significant at the 10% level. We do observe a significant negative effect for retained students on math scores of nine percent of a SD, suggesting that retained students' performance at the top of the distribution might be compromised as retention increases. In contrast, the positive effects on language scores for retained students, and the negative effects for the marginal students, are mostly relevant for middle treated schools. These effects are smaller in magnitude for the high treated group.

This result supports our conjecture about the non-linear nature of the retention effects on test scores. Several conclusions can be obtained from the estimates presented in this section. First, higher retention does not affect math scores, neither at the average nor at any below-median percentiles of the distribution. Second, higher retention positively (negatively) affects language scores for retained students (non-retained students). Third, the fact that the results are stronger for middle treated schools suggests that the effect of retention is non-linear, as these schools obtain larger returns for the same percentage point increase in retention. Higher retention at some point no longer leads to higher scores for the retained students or lower scores for the students at the margin of repeating grade 10. This also shows that our results are not driven by alternative reasons, such as selection. Finally, results obtained for both retained and non-retained students are strongest at the lower end of the test score distribution. This indicates that low ability, retained students might benefit more from increased retention due to repeating classes or because the stigma of retention becomes of lower importance. Non-retained students with similar ability might score lower because they might face a higher threat of being retained, underperforming at the test later on.

Figure 2.4: Effect of the FRP Regime on Test Scores' Distribution



Notes: This figure displays the difference-in-differences effect of the FRP regime on the percentiles of the SABER11 test scores' distribution for both math and language subjects. Panels (a)-(b) present the effect of one and two years prior exposure to the FRP regime for math scores. Panels (c)-(d) plots the same effects for language scores. The caps denote confidence intervals at the 95% significance level.

Table 2.4: Effect of the FRP Regime on Test Scores' Distribution by Multiple Treatment Status

	Math Scores					Language Scores				
	(2) 10th Percentile	(3) 25th Percentile	(4) 50th Percentile	(5) 75th Percentile	(6) 90th Percentile	(8) 10th Percentile	(9) 25th Percentile	(10) 50th Percentile	(11) 75th Percentile	(12) 90th Percentile
$HighTreated_s \times FRP_{t-1}$	0.060 (0.045)	0.039 (0.038)	0.022 (0.036)	0.069* (0.038)	0.082* (0.045)	-0.075 (0.063)	-0.087* (0.051)	-0.095** (0.045)	-0.049 (0.046)	-0.046 (0.053)
$HighTreated_s \times FRP_{t-2}$	0.000 (0.044)	0.012 (0.036)	0.011 (0.034)	-0.006 (0.037)	-0.032 (0.046)	0.140** (0.059)	0.118** (0.046)	0.132*** (0.041)	0.095** (0.042)	0.096* (0.051)
$MiddleTreated_s \times FRP_{t-1}$	-0.017 (0.045)	-0.017 (0.037)	-0.015 (0.035)	-0.007 (0.037)	0.056 (0.046)	-0.153** (0.064)	-0.152*** (0.052)	-0.119*** (0.046)	-0.060 (0.046)	-0.089* (0.054)
$MiddleTreated_s \times FRP_{t-2}$	0.016 (0.044)	0.015 (0.036)	-0.000 (0.034)	-0.013 (0.037)	-0.093** (0.046)	0.187*** (0.060)	0.135*** (0.047)	0.113*** (0.041)	0.074* (0.042)	0.091* (0.051)
$LowTreated_s \times FRP_{t-1}$	0.037 (0.046)	0.020 (0.038)	0.028 (0.036)	0.025 (0.038)	0.050 (0.046)	-0.061 (0.065)	-0.044 (0.054)	-0.085* (0.047)	-0.046 (0.048)	-0.112** (0.055)
$LowTreated_s \times FRP_{t-2}$	-0.001 (0.046)	-0.003 (0.036)	-0.019 (0.035)	-0.009 (0.039)	-0.032 (0.048)	0.089 (0.061)	0.019 (0.048)	0.074* (0.042)	0.050 (0.043)	0.090* (0.052)
Observations	35,693	35,693	35,693	35,693	35,693	35,693	35,693	35,693	35,693	35,693
R-squared	0.022	0.052	0.127	0.223	0.253	0.270	0.163	0.075	0.174	0.299
# Schools	6,248	6,248	6,248	6,248	6,248	6,248	6,248	6,248	6,248	6,248
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors clustered at the school level reported in parentheses. All specifications include fixed effects by school and exam year. Treated schools are defined as in the main text. The outcome variables are schools' different percentiles of the standardised SABER11 test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.6.4 Testing the Common Trend Assumption

In this subsection, we present several robustness checks that support the empirical strategy implemented in this study. Table 2.5 presents difference-in-differences estimates of the common trends specification implied by equation (2.2), using exam year 2007 as a baseline. Appendix Table 2.A1 reports the same analysis for higher, middle, and low treated schools. In both tables we report the F-statistics of the joint test that the AUP-period coefficients are not statistically different from zero.

As our estimates suggest, we can conclude that treated and control schools share a common trend. Regarding the placebo tests implied by equation (2.3), Table 2.6 report difference-in-differences coefficients assuming the FRP regime started in 2008 (Columns (1)-(4)) or 2007 (Columns (5)-(8)). The same estimates considering highly, middle, and low treated schools are reported in Appendix Table 2.A2. As the F-statistics indicate, we do not find evidence that pre-existing trends have a direct impact on the variation in test scores we observe after the retention policy changed.

Finally, we perform a falsification test using the subsample of control schools. From roughly 3000 schools in the control group, we select at random 1500 schools and assign them to the treatment group. Then, we estimate the model implied by our baseline specification (2.1). If we replicate this process say, 1000 times, we should expect to obtain significant results in no more than 50 replications using a 95% confidence level. Otherwise, results from this exercise will cast doubts on our treatment-control classification.

Figure 2.5 displays the absolute t-statistic of each of these replications for our coefficients of interest, where the vertical red line denotes the 5% critical value of a t-student distribution (i.e. 1.96). We also present in Appendix Table 2.A3 the summary statistics of all parameters recovered from this falsification test. As observed, less than 5% of the replications turn out to be significant as only up to 28 replications are statistically different from zero. In addition, all mean coefficients are virtually zero, with standard errors at least 27 times higher than the reported effect. Overall, these results support the treatment-control categorisation used in this chapter.

Table 2.5: Common Trends Assumption Test

	Math Scores		Language Scores	
	(1)	(2)	(3)	(4)
Pre-FRP trends				
$Treated_s \times 1[\text{year}=2008]$	0.011 (0.019)	0.017 (0.019)	0.025 (0.018)	0.029 (0.018)
$Treated_s \times 1[\text{year}=2009]$	-0.021 (0.021)	-0.013 (0.021)	0.030 (0.020)	0.036* (0.020)
$Treated_s \times 1[\text{year}=2010]$	-0.050** (0.024)	-0.038 (0.024)	0.008 (0.019)	0.015 (0.020)
FRP trends				
$Treated_s \times 1[\text{year}=2011]$	-0.014 (0.026)	0.004 (0.026)	-0.045 (0.028)	-0.031 (0.028)
$Treated_s \times 1[\text{year}=2012]$	-0.017 (0.024)	0.004 (0.023)	0.021 (0.019)	0.037* (0.019)
$Treated_s \times 1[\text{year}=2013]$	-0.005 (0.022)	0.018 (0.022)	0.016 (0.019)	0.033* (0.019)
Observations	35,693	35,693	35,693	35,693
R-squared	0.149	0.155	0.094	0.101
# Schools	6,248	6,248	6,248	6,248
Covariates	No	Yes	No	Yes
F-stat (3; 6,248)	2.541	2.043	1.087	1.368
p-value	0.0546	0.106	0.353	0.251

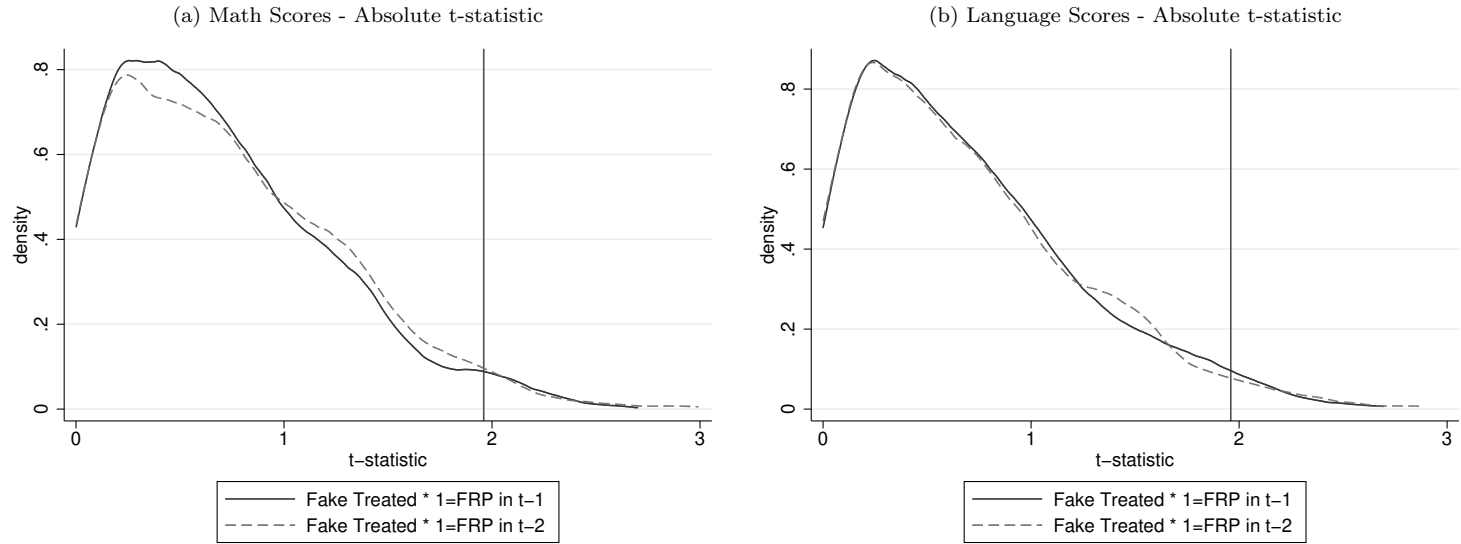
Notes: Robust standard errors clustered at the school level are reported in parentheses. This table shows results for the common trend assumption test. The outcome variables are average standardised test scores for math and language subjects measured for year t . Covariates include the first two lags of: average class size at grade 10, average managerial, health, and support staff per school, average number of teachers under the old and new pay scales, average number of teachers with a professional degree, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. F-statistics reported correspond to the null hypothesis that pre-FRP regime trends differences between control and treated schools are not statistically significant. Treated and controls schools are defined as in the main text. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6: Placebo Tests: FRP Regime Started Before Original Date

	Math Scores		Language Scores		Math Scores		Language Scores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Treated_s \times FakeFRP_{2008,t-1}$	-0.026 (0.018)	-0.021 (0.018)	0.018 (0.017)	0.021 (0.017)				
$Treated_s \times FakeFRP_{2008,t-2}$	-0.001 (0.017)	0.009 (0.017)	-0.030 (0.019)	-0.023 (0.018)				
$Treated_s \times FakeFRP_{2007,t-1}$					0.011 (0.019)	0.018 (0.019)	0.024 (0.018)	0.028 (0.018)
$Treated_s \times FakeFRP_{2007,t-1}$					-0.032* (0.019)	-0.022 (0.018)	-0.018 (0.017)	-0.011 (0.016)
Observations	35,693	35,693	35,693	35,693	35,693	35,693	35,693	35,693
R-squared	0.149	0.155	0.094	0.101	0.149	0.155	0.094	0.101
# Schools	6,248	6,248	6,248	6,248	6,248	6,248	6,248	6,248
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
F-stats (2; 6,248)	1.390	0.692	1.290	0.972	1.520	0.817	0.960	1.233
p-value	0.249	0.500	0.275	0.378	0.219	0.442	0.383	0.291

Notes: Robust standard errors clustered at the school level are reported in parentheses. All specifications include fixed effects by school and exam year. Treated (Control) schools are defined as those with above median (below median) increase in their retention rates at grade 10 from the AUP years to the FRP years. The dependent variables are average standardised math and language SABER11's test scores. Columns (1)-(4) report results on the placebo test assuming the FRP regime started in 2008. Columns (5)-(8) show estimates on the placebo test assuming the FRP regime started in 2007. F-statistics reported correspond to the joint test of the null hypothesis that placebo effects are not different from zero. Covariates include the first two lags of: average class size at grade 10, average managerial, health, and support staff per school, average number of teachers under the old and new pay scales, average number of teachers with a professional degree, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2.5: Falsification Test - 1500 Control Schools as Treated



Notes: Based on 1000 replications. This figure displays absolute t-statistics of the falsification test. We estimate the baseline specification (2.1) using the control group sample only, but we randomly allocate the treatment status to 1500 control schools. All estimations include fixed effects by school and exam year. The outcome variables are mean standardised SABER11 test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of (false) treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. The red vertical line denotes the critical value by which the null hypothesis of non-significance is rejected at the 5% level.

2.6.5 Additional Robustness Checks

As indicated in the data section, there is attrition in our data for schools in which either retention rates or pre-FRP regime’s characteristics are not completely observed throughout the period of interest. To analyse whether this attrition is selective, we report estimates in Table 2.7 using only schools from the seven-years balanced panel. As expected, the size of the estimation sample decreases dramatically, leaving only 3,281 schools left to consider in the estimation. However, we observe that the signs of our estimates do not change. The effects become strongly significant and slightly higher in magnitude, giving strong support to our baseline findings. In fact, given the magnitudes obtained from this robustness check, we can consider the coefficients provided in our baseline results as lower bound estimates of the true effect of the FRP regime on test scores.

Table 2.7: Effect of the FRP Regime on Average Test Scores - Balanced Panel Estimations

	Math Scores		Language Scores	
	(1)	(2)	(4)	(5)
$Treated_s \times FRP_{t-1}$	-0.021 (0.023)	-0.007 (0.022)	-0.091*** (0.032)	-0.079** (0.031)
$Treated_s \times FRP_{t-2}$	0.019 (0.021)	0.024 (0.021)	0.076*** (0.027)	0.079*** (0.027)
Observations	22,967	22,967	22,967	22,967
R-squared	0.176	0.184	0.117	0.126
# Schools	3,281	3,281	3,281	3,281
Covariates	No	yes	No	Yes

Notes: Robust standard errors clustered at the school level reported in parentheses. This table reports difference-in-difference estimates when we consider only those schools from a balanced panel dataset of seven years. All specifications include fixed effects by school and exam year. Treated (Control) schools are defined as those with above median (below median) increase in their retention rates at grade 10 from the AUP years to the FRP years. The outcome variables are average standardised SABER11 test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. Covariates considered in these estimations include the first two lags of: Average class size at grade 10, number of health, support, and managerial non-academic staff per school, number of teachers with a professional degree, number of teachers under the new and old government-regulated pay scales, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Another concern in our empirical strategy is the timing between the announcement of the policy change and the time the new regime was officially in place. As discussed before, schools were informed in 2009 that from 2010 onwards they will be allowed to retain as many students as they prefer. It is plausible then that some schools reacted to this announcement by increasing retention rates in 2009. To check whether our results are robust to this behaviour, we repeat the estimations of our baseline specification, but excluding observations from exam year 2010. Table 2.8 reports difference-in-differences coefficients from this exercise. All coefficients are virtually the same as we obtain in our central findings,

suggesting that schools’ incentives to anticipate the policy change are not the main source of variation driving the effects we are documenting in this chapter.

Table 2.8: Effect of the FRP Regime on Average Test Scores - Excluding Anticipatory Effects

	Math Scores		Language Scores	
	(1)	(2)	(4)	(5)
$Treated_s \times FRP_{t-1}$	-0.012 (0.022)	0.002 (0.022)	-0.068** (0.028)	-0.056** (0.028)
$Treated_s \times FRP_{t-2}$	0.006 (0.019)	0.010 (0.019)	0.066*** (0.023)	0.068*** (0.023)
Observations	30,576	30,576	30,576	30,576
R-squared	0.158	0.166	0.105	0.112
# Schools	6,234	6,234	6,234	6,234
Covariates	No	Yes	No	Yes

Notes: Robust standard errors clustered at the school level reported in parentheses. This table reports difference-in-differences regressions excluding observations from exam year 2010. All specifications include fixed effects by school and exam year. Treated (Control) schools are defined as those with above median (below median) increase in their retention rates at grade 10 from the AUP years to the FRP years. The outcome variables are average standardised SABER11 test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. Covariates considered in these estimations include the first two lags of: Average class size at grade 10, number of health, support, and managerial non-academic staff per school, number of teachers with a professional degree, number of teachers under the new and old government-regulated pay scales, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

2.7 Potential Mechanisms

In this section we explore propagation channels that may drive the effects we obtain. As implied by the FRP regime, treated schools significantly increased their retention rates, relative to schools in the control group. Are there any school characteristics that induce some institutions to retain more students? Are some school attributes amplifying the impacts of increased retention? To answer these questions, we assess the extent to which average class size at grade 10, teachers’ qualifications, and changes in the way teachers are remunerated play a role in disseminating the effects of grade retention.

There is a large consensus in the economics of education literature about the negative effects of large class sizes on students’ academic performance (Angrist and Lavy, 1999; Fredriksson, Öckert, and Oosterbeek, 2012). Nonetheless, to our knowledge there is no discussion on whether grade retention and class size at school exhibit some complementarities. Assuming everything else constant, increased retention may have a positive impact on class size. We can also reverse the direction of the relationship. Schools with more students per group might have fewer incentives to retain students as large classrooms are more difficult to

manage. Hence, we might expect the positive (negative) effects of retention to be weaker (stronger) on retained (marginal) students as the number of pupils per group rises.

Regarding our second transmission channel, recent papers highlight the empirical challenges of identifying the effects of teacher quality in the classroom (Gerritsen, Plug, and Webbink, 2017; Rivkin, Hanushek, and Kain, 2005). We may expect the benefits (costs) of retention to be amplified (reduced) as teachers' education improves.

We exploit a regulation change in the way public school teachers are remunerated. From 2002 onwards, under Decree 1278, the remuneration, probation period, and screening process for newly hired teachers changed substantially. Under the new system, prospective teachers need to participate in a public entry contest which, after completion, will determine their starting rank and wage. In addition, teachers hired under this new scheme will be subject to a probation period up to 12 months, to then receive tenure that can be revoked if subsequent performance evaluations are not satisfactory. In contrast, teachers hired before June 2002 were subject to the old 1979's, more lenient regulation (Decree 2277). This innovation in the employment relationship of teachers created a mixture of academic staff paid with the old and new pay scales. As it is expected that newly hired teachers will replace those about to retire, the proportion of teachers under the new pay scale at school is a key amplifying mechanism to study. A priori, the direction of the effect is unclear. There is empirical evidence suggesting that teachers under the new regulation are positively selected, implying positive but moderate effects on school performance (Brutti and Sánchez, 2017). As career concerns are prevalent in their probation period, teachers might have incentives to put more effort. Thus, increasing participation of teachers under the new pay scale might have positive effects on students, especially those at the lower end of the ability distribution.

To test these mechanisms, we modify our baseline specification as follows:

$$\begin{aligned}
 Y_{st} &= \alpha_s + \delta_t + \sum_{h=1}^2 \gamma_h [Treated_s \times FRP_{t-h}] + \sum_{h=1}^2 \beta_h Attribute_{t-h} \\
 &+ \sum_{h=1}^2 \rho_h [Treated_s \times FRP_{t-h} \times Attribute_{t-h}] + \varepsilon_{st}, \tag{2.4}
 \end{aligned}$$

where the variable $Attribute_{t-h}$ denotes each mechanism we intend to test, one and two years before the exam takes place. The coefficients of interest in this specification are ρ_1 and ρ_2 , which measure how each attribute in question propagates the effects of increased retention for marginal and retained students, respectively. We present our findings from this analysis in Table 2.9. Panel A reports the difference-in-differences coefficients. Panels B, C, and D show estimates on the interaction of the difference-in-differences effects with average school's class size at grade 10, the proportion of teachers with a post-secondary education degree, and the proportion of teachers under the new, government regulated pay scale, respectively.

Table 2.9: Effect of the FRP Regimen on SABER11 Test Scores: Mechanisms

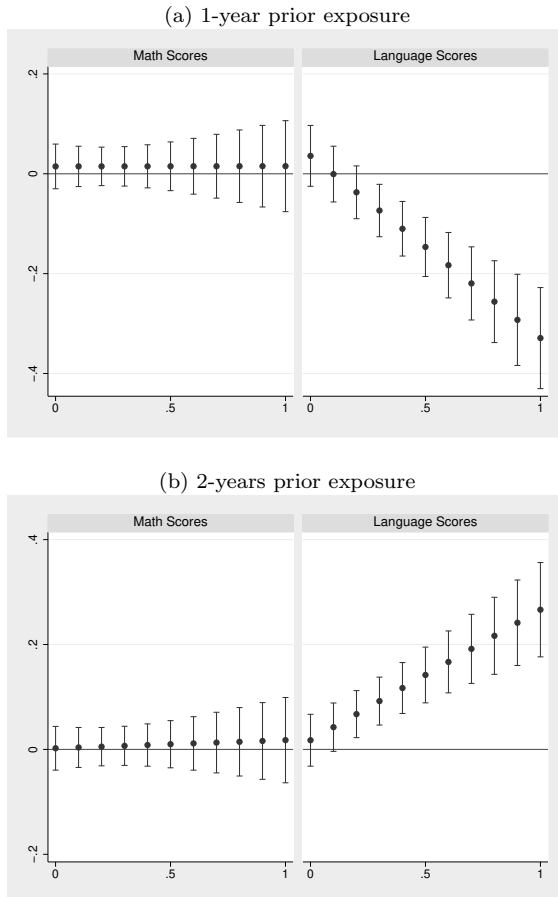
	Math Scores			Language Scores		
	(1)	(2)	(3)	(4)	(5)	(6)
<hr/> Panel A: Difference in Differences Effect <hr/>						
$DID_{s,t-1} = Treated_s \times FRP_{t-1}$	0.006 (0.029)	0.073 (0.067)	0.015 (0.023)	-0.079** (0.036)	0.063 (0.078)	0.036 (0.031)
$DID_{s,t-2} = Treated_s \times FRP_{t-2}$	0.036 (0.027)	-0.006 (0.058)	0.002 (0.021)	0.051* (0.030)	0.155*** (0.056)	0.017 (0.025)
<hr/> Panel B: Effects by Class Size <hr/>						
$DID_{s,t-1} \times ClassSize_{s,t-1}$	0.000 (0.001)			0.001 (0.001)		
$DID_{s,t-2} \times ClassSize_{s,t-2}$	-0.001* (0.001)			0.000 (0.000)		
<hr/> Panel C: Effects by Teacher's Qualifications <hr/>						
$DID_{s,t-1} \times TeachersQual_{s,t-1}$		-0.064 (0.068)			-0.125 (0.077)	
$DID_{s,t-2} \times TeachersQual_{s,t-2}$		0.015 (0.058)			-0.093* (0.054)	
<hr/> Panel D: Effects by Teacher's Pay Scale <hr/>						
$DID_{s,t-1} \times TeachersNewpay_{s,t-1}$			0.001 (0.054)			-0.365*** (0.060)
$DID_{s,t-2} \times TeachersNewpay_{s,t-2}$			0.015 (0.047)			0.249*** (0.050)
Observations	35,693	35,693	35,693	35,693	35,693	35,693
R-squared	0.155	0.155	0.155	0.101	0.102	0.104
# Schools	6,248	6,248	6,248	6,248	6,248	6,248

Notes: Robust standard errors clustered at the school level reported in parentheses. This table presents difference-in-differences estimates and their interactions with different school attributes. All specifications include fixed effects by school and exam year. Treated (Control) schools are defined as those with above median (below median) increase in their retention rates at grade 10 from the AUP years to the FRP years. The outcome variables are average standardised SABER11 test scores for math and language subjects at year t . *ClassSize* measures the average number of students per group at grade 10. *TeachersQual* accounts for the proportion of teachers with a post-secondary education degree per school. *TeachersNewpay* measures the proportion of teachers per school under the new pay scale regulated by the central government. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

As observed in Columns (1)-(3) none of the mechanisms considered is masking the null effect of grade retention on math scores. We obtain a marginally significant effect of class size for retained students, but we claim that this can be ignored as it is only significant at the 10% level and very small in magnitude. In contrast, results for language scores indicate two propagation mechanisms worth to be discussed.

First, we observe a negative effect of increased teachers' qualifications on test scores, as both coefficients of interest exhibit a negative sign. However, it seems that the effect in question is relevant (at the 10% significance level) only for retained students. These findings support the idea that policy interventions aimed to foster the human capital acquisition of teachers may not be as effective as other measures to extract the largest gains from grade retention.

Figure 2.6: Mechanisms: Variation in Teachers' Composition.



Notes: This figure plots the marginal effects from a difference-in-differences estimation, interacted with the proportion of teachers under the new pay scale. each cap denotes confidence intervals at the 95% level of a 10-percentage points increase in the proportion of teachers under the new pay scale.

Second, we observe the teachers' composition at treated schools to play a key role in propagating the effects of increased retention. In particular, a 10% jump in the proportion of newly hired teachers implies a rise (drop) in language scores of 2.5% (3.6%) of a SD for retained (non-retained) students. To present these effects in more detail, in Figure 2.6 we plot the marginal effects of a 10 percentage points increment in the proportion of teachers hired under the new pay scale, one and two years before the exam takes place (panels (a) and (b), respectively). As observed, it is clear that benefits and costs of increased retention for language scores are monotonically increasing. This fact suggests that retained students benefit more from a relatively young academic workforce that is willing to invest time and effort in their education.

2.8 Concluding Remarks

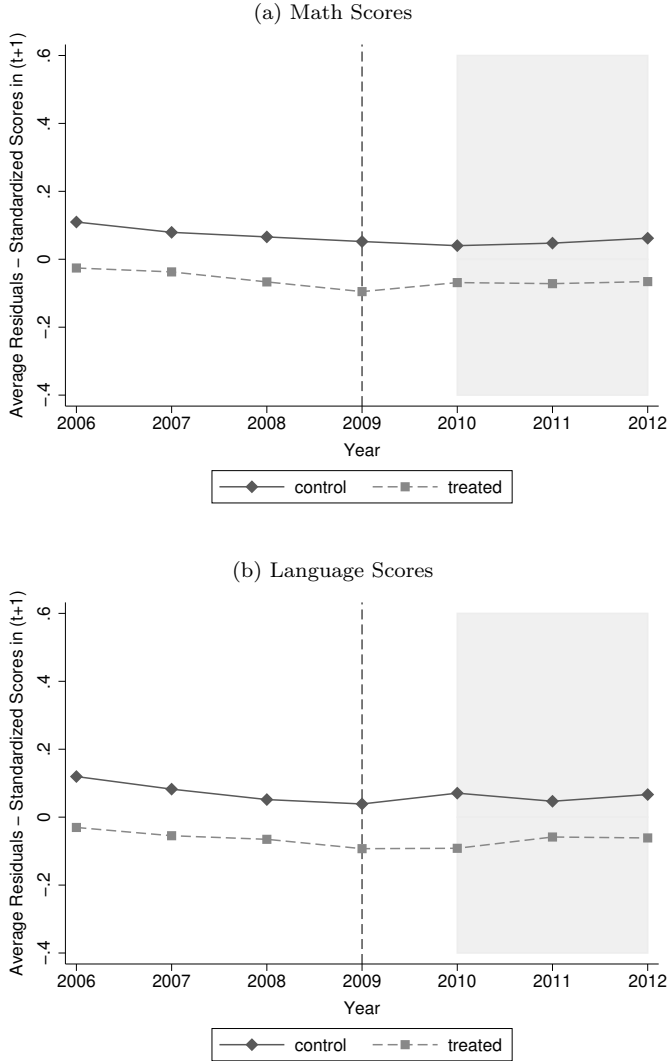
In this chapter, we have analysed the effect of retention in grade 10 on school performance in grade 11. We exploit a law change in Colombia with respect to retention. Until 2010, schools were allowed to retain a maximum of 5 percent of their total number of students. After the abolishment of the law in 2010, schools were free to retain as many students as they considered appropriate. This led to a large increase in retention, with considerable heterogeneity across schools. We use a difference-in-differences analysis to study the effect of retention on test performance. Placebo tests suggest that there are common trends in scores among schools that responded in various degrees to the law change.

Our estimates reveal that there are positive effects of more retention on language test performance for retained students. These effects are non-linear, as modest increases in retention have positive effects but larger increases in retention do not necessarily lead to better performance. In addition, our findings suggest that non-retained students at the lower end of the ability distribution perform worse in language tests. Potential explanations for this effect include the negative spillover impacts from formerly retained students, the strategic substitution of effort between stem and non-stem subjects in order to avoid grade retention, and the changes in teachers' workforce composition. In contrast, we do not find any effects on math scores that can be attributed either to retained or non-retained pupils.

This research shows the importance of analysing effects of retention at different margins of the ability distribution. Although data restrictions do not allow to recover information on individual retention, we feel confident that the empirical strategy and data construction implemented in this chapter aids to solve this limitation by decomposing the effect of retention among different types of students. More research is needed to investigate whether the gains of retention we identify can be outweighed by other costs of retention, such as school dropouts, career choice regret, delayed (or sudden) labour market participation, forgone income, and the formation of undesirable personality traits, preferences and risk attitudes across the life cycle.

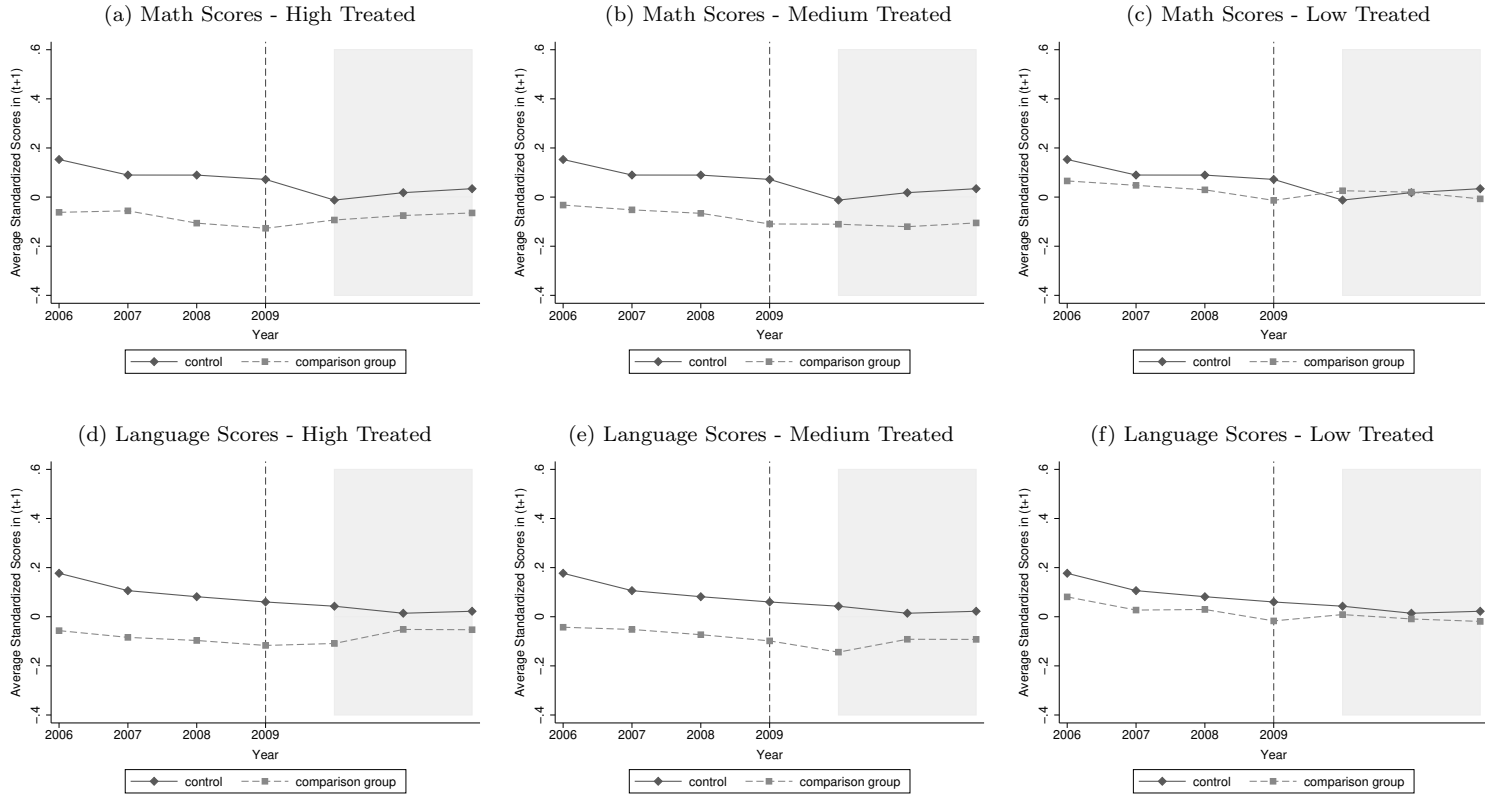
Appendices to Chapter 2

Figure 2.A1: Test Scores Residuals by Treatment Status.



Notes: This figure displays average residuals from an OLS regression where the dependent variable is the average test scores in year $(t+1)$ as a function of school and exam-year fixed effects, and a set of school specific covariates. Panel (a) presents common trends on math scores' average residuals between treated (dashed lines) and control (solid lines) schools. Panel (b) shows similar trends on language test scores' average residuals. Covariates considered in these estimations include the first two lags of: Average class size at grade 10, number of health, support, and managerial non-academic staff per school, number of teachers with a professional degree, number of teachers under the new and old government-regulated pay scales, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree.

Figure 2.A2: Test Scores Residuals by Treatment Status - Multiple Treatment Groups



Notes: This figure shows average residuals from an OLS regression where the dependent variable is the average tests scores in year $(t + 1)$ as a function of school and exam-year fixed effects, and a set of schools' specific covariates. Panels (a)-(c) show math scores' average residuals. Panels (d)-(f) present the results for language scores. The dashed vertical line denotes year 2009 when schools were informed that the AUP regime will no longer hold. The gray area denotes the years when the FRP regime was in place. Covariates include the first two lags of: Average class size at grade 10, number of health, support, and managerial non-academic staff per school, number of teachers with a professional degree, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree.

Table 2.A1: Common Trend Assumption Test by Heterogeneous Treatment

	Math Scores		Language Scores	
	(1)	(2)	(3)	(4)
$HighTreated_s \times 1[\text{year}=2008]$	0.034 (0.030)	0.044 (0.030)	0.032 (0.030)	0.040 (0.030)
$HighTreated_s \times 1[\text{year}=2009]$	-0.021 (0.033)	-0.008 (0.033)	0.046 (0.031)	0.057* (0.031)
$HighTreated_s \times 1[\text{year}=2010]$	-0.047 (0.037)	-0.029 (0.037)	0.027 (0.031)	0.039 (0.031)
$MiddleTreated_s \times 1[\text{year}=2008]$	0.003 (0.032)	0.011 (0.032)	0.032 (0.031)	0.038 (0.031)
$MiddleTreated_s \times 1[\text{year}=2009]$	-0.028 (0.035)	-0.018 (0.035)	0.029 (0.033)	0.038 (0.033)
$MiddleTreated_s \times 1[\text{year}=2010]$	-0.080** (0.039)	-0.066* (0.039)	-0.002 (0.032)	0.008 (0.032)
$LowTreated_s \times 1[\text{year}=2008]$	0.000 (0.031)	0.006 (0.031)	-0.025 (0.030)	-0.021 (0.031)
$LowTreated_s \times 1[\text{year}=2009]$	-0.013 (0.035)	-0.003 (0.035)	0.024 (0.032)	0.031 (0.033)
$LowTreated_s \times 1[\text{year}=2010]$	-0.047 (0.037)	-0.035 (0.037)	-0.012 (0.032)	-0.002 (0.033)
Observations	35,693	35,693	35,693	35,693
R-squared	0.150	0.156	0.096	0.103
# Schools	6,248	6,248	6,248	6,248
Covariates	No	Yes	No	Yes
F-stat (9 , 6614)	1.051	0.968	1.064	1.220
p-value	0.397	0.464	0.386	0.277

Notes: Robust standard errors clustered at the school level are reported in parentheses. This table shows results for the common trend assumption test. All specifications include fixed effects by school and exam year. For matters of space, we only report the coefficients from the pre-FRP trends years. The outcome variables are average standardised test scores for math and language subjects measured for year t . Covariates include the first two lags of: average class size at grade 10, average managerial, health, and support staff per school, average number of teachers under the old and new pay scales, average number of teachers with a professional degree, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. F-statistics reported correspond to the null hypothesis that pre-FRP regime trends differences between control and treated schools are not statistically significant. Treated and controls schools are defined as in the main text. *** p-value < 0.01, ** p-value < 0.05. * p-value < 0.1.

Table 2.A2: Placebo Test Using Multiple Treatment Groups: FRP Regime Started Before 2010

	Math Scores		Language Scores		Math Scores		Language Scores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: FRP started in 2008								
$HighTreated_s \times FakeFRP_{2008,t-1}$	-0.037 (0.028)	-0.029 (0.028)	0.030 (0.026)	0.037 (0.026)				
$HighTreated_s \times FakeFRP_{2008,t-2}$	0.025 (0.028)	0.038 (0.028)	-0.027 (0.029)	-0.018 (0.028)				
$MiddleTreated_s \times FakeFRP_{2008,t-1}$	-0.029 (0.029)	-0.023 (0.029)	0.013 (0.027)	0.018 (0.027)				
$MiddleTreated_s \times FakeFRP_{2008,t-2}$	-0.034 (0.027)	-0.024 (0.027)	-0.058** (0.029)	-0.051* (0.029)				
$LowTreated_s \times FakeFRP_{2008,t-1}$	-0.012 (0.030)	-0.006 (0.030)	0.036 (0.027)	0.041 (0.027)				
$LowTreated_s \times FakeFRP_{2008,t-2}$	-0.004 (0.028)	0.001 (0.028)	-0.056* (0.030)	-0.053* (0.030)				
Panel B: FRP started in 2007								
$HighTreated_s \times FakeFRP_{2007,t-1}$					0.035 (0.030)	0.046 (0.030)	0.031 (0.030)	0.040 (0.030)
$HighTreated_s \times FakeFRP_{2007,t-2}$					-0.035 (0.030)	-0.022 (0.030)	-0.007 (0.027)	0.002 (0.026)
$MiddleTreated_s \times FakeFRP_{2007,t-1}$					0.003 (0.032)	0.010 (0.032)	0.031 (0.031)	0.037 (0.031)
$MiddleTreated_s \times FakeFRP_{2007,t-2}$					-0.057* (0.030)	-0.047 (0.030)	-0.049* (0.027)	-0.041 (0.026)
$LowTreated_s \times FakeFRP_{2007,t-1}$					0.001 (0.031)	0.006 (0.031)	-0.026 (0.030)	-0.022 (0.031)
$LowTreated_s \times FakeFRP_{2007,t-2}$					-0.016 (0.031)	-0.008 (0.031)	0.004 (0.027)	0.010 (0.027)
Observations	35,693	35,693	35,693	35,693	35,693	35,693	35,693	35,693
R-squared	0.149	0.155	0.094	0.101	0.149	0.155	0.094	0.101
# Schools	6,248	6,248	6,248	6,248	6,248	6,248	6,248	6,248
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
F-stats (6; 6,614)	1.535	1.462	1.456	1.478	1.181	1.146	1.808	2.015
p-value	0.162	0.187	0.189	0.181	0.313	0.333	0.0934	0.0602

Notes: Robust standard errors clustered at the school level are reported in parentheses. This table presents results on placebo tests for the baseline difference-in-differences specification, assuming the FRP started in year 2008. All specifications include fixed effects by school and exam year. Treated (Control) schools are defined as those with above median (below median) increase in their retention rates at grade 10 from the AUP years to the FRP years. The dependent variables are average standardised math and language SABER11's test scores. F-statistics reported correspond to the joint test of the null hypothesis that placebo effects are not different from zero. Covariates include the first two lags of: average class size at grade 10, average managerial, health, and support staff per school, average number of teachers under the old and new pay scales, average number of teachers with a professional degree, proportion of teachers under the new pay scale, and proportion of teachers with a professional degree. *** p-value < 0.01, ** p-value < 0.05. * p-value < 0.1.

Table 2.A3: Falsification Test - 1500 Control Schools as Treated

	Math Scores				Language Scores			
	(1) mean	(2) std. dev.	(3) min	(4) max	(5) mean	(6) std. dev.	(7) min	(8) max
<hr/>								
$\underline{Treated_s \times FRP_{t-1}}$								
Coefficient	0.001	0.026	-0.067	0.079	0.001	0.035	-0.108	0.100
Std. Error	0.029	0.000	0.029	0.029	0.040	0.000	0.040	0.040
t-statistic	0.705	0.516	0.001	2.700	0.696	0.529	0.000	2.694
$P(t > T_{5\%})$	0.024	0.153	0.000	1.000	0.027	0.162	0.000	1.000
<hr/>								
$\underline{Treated_s \times FRP_{t-2}}$								
Coefficient	-0.001	0.025	-0.081	0.069	-0.000	0.031	-0.100	0.095
Std. Error	0.027	0.000	0.027	0.027	0.035	0.000	0.035	0.035
t-statistic	0.741	0.539	0.001	2.992	0.698	0.536	0.000	2.866
$P(t > T_{5\%})$	0.026	0.159	0.000	1.000	0.028	0.165	0.000	1.000
# Replications	1000	1000	1000	1000	1000	1000	1000	1000

Notes: Based on 1000 replications. This table reports difference-in-difference estimates on the control group sample when we randomly allocate treatment status to 1500 control schools. All specifications include fixed effects by school and exam year. The outcome variables are average standardised test scores for math and language subjects at year t . The coefficients of interest are the interaction of an indicator of (false) treatment status with a set of dummy variables FRP_{t-1} and FRP_{t-2} , measuring the exposure to the FRP regime one and two years before the SABER11 exam is taken, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The effect of grade retention on secondary school dropout:
Evidence from a natural experiment

Abstract

This chapter analyses the effects of grade retention on secondary school dropout by evaluating a retention policy reform introduced in 2010 in Colombia, which ended the restriction that the annual number of retained students at a school could not exceed 5 percent of the total school population. Using administrative data at the school level, we estimate a difference-in-differences model that exploits variation in schools' retention rates before and after the reform. We distinguish dropout rates by grade (grade 6 to 11). Moreover, we distinguish between retained students who dropped out of school by the end of the year of their retention and the dropout effect on all students enrolled in school the year after retention. Our robust estimates reveal that higher retention increases the rate of students dropping out of school the same year of their retention, that means without enrolling to repeat the failed grade. However, there is little, if any, causal effect of grade retention on the dropout rates of all other students enrolled in the school one year after retention. We find that the latter effect is stronger when retention takes place at the earlier grades whereas the effect for retained students only is strongest when retention occurs at grade 9 and grade 11, when students would be entitled to receive the lower secondary school certificate and the high-school diploma respectively.

JEL Classification: I2.

I am grateful to Lex Borghans, Bart Golsteyn, Andries de Grip, Annemarie Künn-Nelen, Olivier Marie, and Sergio Parra-Cely for their insightful comments and suggestions. I also thank the Colombian Bureau of Statistics (DANE) for making available their datasets on the C-600 census.

3.1 Introduction

Despite increasing attention by policy makers, school dropout is still a serious issue of particular importance in developing countries, where universalisation of education continues to be a challenge.¹ Although graduation from secondary school is considered the minimum level of educational attainment needed for successful participation in further education and the labour market, the numbers of young children leaving school without completing a secondary qualification are quite large in some countries (Lamb and Markussen, 2011; OECD, 2012).²

Both in public policy and in the academic literature, grade repetition is often associated to low enrolment, low graduation rates and high dropout rates (e.g. Bowers et al., 2012; Busso, Bassi, and Muñoz, 2013; Ikeda and García, 2014; Jimerson et al., 2002; OECD, 2016; Roderick, 1994; Stearns et al., 2007). In the literature, it is commonly conjectured that being retained may influence factors associated with dropping out of school such as low self-esteem, socio-emotional adjustment, peer relations, and school engagement.

An important caveat in the literature is that the largest part of the evidence remains strictly correlational. Although correlations are informative, if students are selected into retention on the basis of unobservable factors, important confounders may lead to biased conclusions regarding the actual effects of retention on school dropout. In addition, the possibility that the strength of the association between retention and dropout may differ depending on the timing of retention during the secondary school cycle has been largely overlooked. These existing research gaps are most notable in the context of developing countries, where most of the studies are restricted to trend analyses (e.g. Busso et al., 2013; Di Gropello, 2006).

In this study, we contribute to filling these gaps in three ways. First, we propose a difference-in-differences framework to find the causal effect of retention on secondary school dropout rates, distinguishing between (1) retained students who drop out of school by the end of the year of their retention and (2) the dropout effect on all students enrolled in school the year after retention. Second, we analyse the heterogeneity of these effects at different moments of the education cycle by estimating the impact of retention on dropouts at each grade of secondary school from grade 6 to 11. Last, we are able to examine the heterogeneity and (non)linearity of the main effects with regard to treatment intensity by distinguishing between students in schools that reacted differently to the policy reform from which we obtain the plausibly exogenous variation in school retention rates.

¹ School dropout is also a central theme in the policy agendas of developed countries. For instance, the European Horizon 2020 Strategy and the US No Child Left Behind Act include policy targets to reduce early school-leaving.

² Nearly 40 percent of adolescents (between 15 and 19 years old) in Latin America drop out of school before completing the secondary school cycle. The greatest dropout rates occur at upper secondary education. Most students leave school during the first year of secondary education, one of the critical points for dropping out. Approximately 45 percent of those that enter secondary education do not graduate. Upper secondary school completion rates in Latin America are well below the average for OECD countries, of around 85 percent (Kattan and Székely, 2015; OECD, 2011).

To overcome the endogeneity problem, we exploit a policy change with respect to retention in Colombia. Until 2009, schools were restricted by national regulation to retain up to a maximum of 5 percent of their students. This retention restriction was ended by the Ministry of Education through a new regulation mandate, allowing schools from 2010 onwards to retain as many students as they considered necessary, and thereby giving them more discretion in their promotion approaches. We use the term *Free Retention Policy* (FRP) reform to refer to this policy change.

To estimate the effects of grade retention on school dropout rates, we use administrative school-grade records across nine school years and exploit variation in retention rates at the school level, as induced by the FRP reform in Colombia. This reform increased retention rates dramatically in some schools, while in others it had no effect.

Since similar schools reacted differently to the policy change, our approach compares the grade dropout rates in secondary school before and after the reform, between schools where retention rates increased a lot and schools where retention rates remained relatively constant across both periods. Treatment and control groups are defined by the above-median historical increase in retention attributed to the law change, analogue to the method used in recent papers in a different context than ours (Bauernschuster et al., 2016; Havnes and Mogstad, 2011). Schools in which retention rates increased more than the median change are labelled the treatment group and those that responded less than the median, the control group. Several placebo and falsification tests show that pre-treatment trends in dropout rates were similar in treatment and control schools, indicating that the key identifying assumption of our difference-in-differences model holds and, therefore, we can confirm that the estimated effects on dropout rates can be attributed to the FRP reform only.

We distinguish between two outcomes as the effects of the reform. First, we consider *end-of-year dropout rates*, referring to the proportion of pupils from a cohort enrolled in a given grade, who concluded the school year but were retained, and did not register in the educational system to repeat the failed grade. Since this is the rate of retained students who stop their school education in the same year of retention, we measure end-of-year dropout rates in year t as the outcome of retention rates during the same year t . Second, we consider *early dropout rates* that refer to the proportion of all students enrolled in each grade who abandoned the educational system without concluding the school year. This is the dropout rate among all students enrolled in year $t + 1$, which gives insights into the spill-over effects of retention on non-retained students. Therefore, early dropout rates in year t are measured as the outcome of retention rates in year $t - 1$.

We highlight two major findings. First, a remarkably large positive effect of grade retention on end-of-year dropout rates among retained students and a positive but relatively small effect of grade failure on consecutive early dropout rates among all students enrolled in education. Second, we show significant heterogeneity in both effects depending on the timing of retention along the secondary education cycle: the effects of grade failure on early dropout rates are stronger when retention takes place at the earlier grades whereas the effect on

retained students is strongest if retention occurs at grade 9 and grade 11. These are precisely the grades for which successful completion entitles the students to receive the lower secondary school certificate and the high-school diploma, respectively.

We first provide evidence of a significant positive effect of increased retention – as induced by the FRP reform – on the average dropout rate of secondary school students. The effect of increased retention on early dropout seems to be rather small: on average, a 10 percentage points rise in retention rates increases early dropout rates by 0.8 of a percentage point in the year following retention. However, the same increase in retention of 10 percentage points is shown to cause an increase of 4.2 percentage points in end-of-year dropout rates for retained students, which means that approximately 40 percent of retained students after the reform did not continue with their secondary education, after learning they needed to repeat the grade. Both our main effects seem to be linearly proportional to the extent of retention growth, that is, the strongest effect on dropout rates was experienced by students enrolled in schools where retention rates increased the most.

Second, we show that the overall positive effect of retention on early dropout rates is stronger in the earlier grades of secondary school and that there is not any significant effect during the last two years of high school, i.e. grades 10 and 11. This means that the strongest effect we observe occurs at grade 6, the year of transition from primary to secondary education. Conversely, the retention effect on the dropout rates of retained students is not only significant across all six grades but strongest if students are retained at the end of grade 9 and grade 11, suggesting there is a high cost of retention at these grades due to the forgone opportunity for students to finalise either one of the two cycles of secondary education.

Finally, we observe further heterogeneity in the effect of retention across the six grades of secondary education if we distinguish between schools treated with different intensity. Although we find most of our main effects to be linear to schools' retention growth, we observe that retained students from grade 9 to grade 11 in mid-treated schools show higher dropout rates than similar students in high-treated schools where retention rates were even higher. This last finding suggests that students retained at later stages of secondary education in a school environment where retention is not common can cause more harm to the future perspectives of those students than if they would have been in schools where retention is more pronounced or a more likely to occur at the end of the school year.

The remainder of the chapter is organised as follows. Section 3.2 discusses the literature related to our research question. Section 3.3 summarises the Colombian educational system and the FRP reform we exploit. Section 3.4 presents our model and empirical strategy. Section 3.5 describes our data and descriptive statistics. In Section 3.6 we discuss the plausibility of the identifying assumptions and present our empirical results and robustness analyses. Section 3.7 concludes the chapter.

3.2 Related Literature

Our study contributes to the literature in various disciplines that have analysed whether – and to what extent – retention determines later school completion and affects the probabilities of dropping out of school.

There is extensive educational research on the determinants of high school dropout, largely based on survey data. Many of these studies find numerous individual characteristics associated with above average rates of dropout. Socioeconomic status (usually measured by parental education, occupational status, or income) is one of the most consistent explanatory factors (Alexander and Kabbani, 2001; Rumberger, 1995, 2004). Moreover, different generations of migrant children and male students (Pharris-Ciurej, Hirschman, and Willhoft, 2012) and adolescents from single parent families and those that experience more residential mobility are shown to have a higher risk of dropping out of school (McLanahan and Sandefur, 1994; Rumberger and Larson, 1998). School-related characteristics are revealed as determinants of dropout over and above individual and family-related motives (Dalton, Glennie, and Ingels, 2009; Rumberger, 2004). Among these and several other factors that explain early school leaving, grade retention is often referred to as a strong predictor of school dropout.

Studies on dropouts consistently find that repeating a grade is positively associated with leaving school before graduation (e.g. Bowers, 2010; Bowers et al., 2012; De Witte, Cabus, Thyssen, Groot, and van den Brink, 2013; Janosz, LeBlanc, Boulerice, and Tremblay, 1997; Jimerson, 1999, 2001; Jimerson et al., 2002; Plank and A., 2005; Roderick, 1994; Rumberger and Larson, 1998; Stearns et al., 2007; Temple, Reynolds, and Ou, 2004). One of these studies (Stearns et al., 2007) illustrates that various school resources affect differently the association between elementary school retention and high school dropouts for white, black, and Latino students in the US.

Regarding the possible explanations for the positive association between retention and dropouts, some scholars have conjectured that it may reflect: (1) the difficulty that over-age retained students may face in integrating themselves into peer and school cultures (Plank and A., 2005; Roderick, 1994), (2) the self-esteem frustration associated with the failure sentiment, the stigma of being unintelligent, and lagging behind, which can increase the likelihood of leaving school permanently, rather than just temporarily (Alexander and Kabbani, 2001; Plank and A., 2005), (3) the psychological and behavioural disengagement from school due to loss of social capital that students draw from their relationships with teachers, parents, and peers (Alexander and Kabbani, 2001; Lamote, Speybroeck, Van Den Noortgate, and Van Damme, 2013; Stearns and Glennie, 2006), (4) the (subconscious) response of teachers to the retention status, which may make them expect and demand less school achievement from retained students and put less effort into reaching them (De Witte et al., 2013; Jimerson et al., 2002; Stearns et al., 2007), and (5) the long-term trajectory of low grades or poor school performance (Bowers and Sprott, 2012).

An important caveat is that these studies can only report correlations and not causal estimates. Since practical constraints make retention experimental designs difficult to implement, most of these studies can only conduct statistic

association analyses or build structural models to test the effect of grade retention on dropouts while controlling for other covariates, especially achievement. Although correlations are informative, important confounders may bias such estimates. Retention is defined very differently across the studies, e.g. identifying over-age students for their grade, asking students if they have ever been retained, examining retention school records, or restricting a definition of retention to specific grade levels. This leads to differences in precision and specificity across the studies due to the retention definition, as well as which grade level was included in the definition and how dropout was defined in each study. A second caveat is that researchers have not assessed the possibility that the strength of the association between retention and dropout may differ, depending on the timing of retention. This is an important problem, given suggestive evidence that various predictors of dropping out vary in strength and significance at different stages of high school but also evidence that retention has a heterogeneous effect on the school performance of students, depending on the timing of retention and the time since retention elapses (Fruehwirth et al., 2016; Stearns and Glennie, 2006).

Last, it is worth noting the existing gap in this research especially in the context of developing countries, where dropping out of school is an even more pronounced phenomenon. Related studies in the Latin-American context often look at trends in dropout but not the reasons behind it. The literature does not offer explanations to the dropout phenomenon explicitly, nor to the trends observed during the past decades (e.g. Busso et al., 2013; Cabrol, 2002; Di Gropello, 2006; Ikeda and García, 2014).

Three articles in the economics literature are closest to our study. First, Eide and Showalter (2001) use an instrumental variable for retention, based on exogenous variation across states in kindergarten entry dates, to evaluate the effect of retention on students' dropout rates and labour market earnings. Their results for white students suggest that grade retention may have some benefit to students by lowering dropout rates; however, their IV estimates are statistically indistinguishable from zero. For black students, the IV approach gave very poor first-stage results. Their own conclusion is that the IV estimates are not conclusive neither informative enough.

Second, Jacob and Lefgren (2009) use plausibly exogenous variation in retention – at grades 6 and 8 – generated by a test-based promotion policy in Chicago to assess the long-run effects of retention on high school completion. They find that retention among younger students (grade 6) does not affect the likelihood of high school completion but retaining low-achieving eighth grade students in elementary school increases the probability that these students will drop out of high school. According to their study, it appears that the differential effect across grade level is driven by the fact that students retained in earlier grades have more opportunities to catch up with their peers and, conversely, students who narrowly missed retention in earlier grades have more opportunities to 'fall back' in subsequent years. The main difference of our study with Jacob and Lefgren's (2009) research is that instead of a regression discontinuity design that relies on the specific margin at which retention was increased by the new tests thresholds for students to be promoted to the next grade, we can analyse the

dropout rates at several moments of the distribution of students retention, at all grades, since the policy change we exploit affected treated students heterogeneously in the full range of subjects not conditioned to changes in any test or achievement results, which we believe remained unaffected before and after the reform.

Third, Manacorda (2012) studies the effects of retention in secondary schools (grades 7 to 9) in Uruguay on dropout rates and school attainment, exploiting a discontinuity established by a rule of automatic grade failure for pupils with more than three failed subjects at the end of the school year. Using administrative longitudinal microdata, this analysis reveals that grade failure induces students to drop out at the end of the school year when failure occurs, which has long-lasting negative effects on their school attainment. Apart from the same advantage of not restricting our analysis only to the specific margin of the discontinuity, as mentioned earlier, we expand the results of Manacorda's (2012) article in two more ways. First, we do not face the discontinuity-related concern that assignment around the threshold of three failed subjects might not be as good as random due to the strategic behaviour of better-performing students or possible manipulation of final scores by teachers to promote students with better latent outcomes. Second, we have several years of information before and after the retention reform we exploit, whereas Manacorda (2012) only has two years of data on failed subjects, information that is not available at the end of the school year but only at an intermediate term. This retention measure is not precise since students with three or fewer failed subjects by the intermediate term might have eventually failed and some students with four or more failed subjects at their mid-term could have eventually passed. Our administrative source of data contains actual students' retention results by the end of each grade in each school cycle.

Finally, it is worth to briefly mention that the growing literature on early school leaving indicates that dropout has profound social and economic consequences for students, their families, and their communities. Several studies suggest that school dropouts, compared with their graduated peers, are more frequently associated with higher rates of and long-term unemployment, poverty, bleak health prospects, sustained dependence on public assistance, single parenthood (in females), political and social apathy, (juvenile) crime and lower overall lifetime earnings and life expectancy (e.g. Bowers et al., 2012; De Witte et al., 2013; Psacharopoulos, 2007; Rumberger, 2011; Rumberger and Lamb, 2003; Swanson, 2004).

3.3 Background

3.3.1 The Colombian Educational System

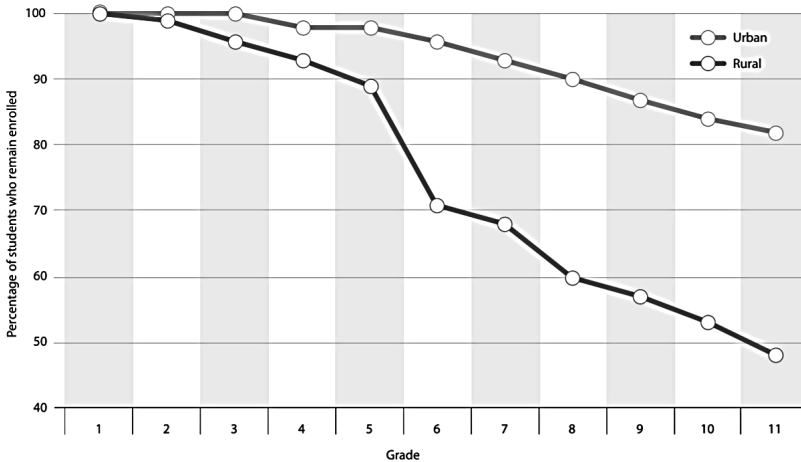
Colombia has an eleven-year system of elementary and secondary education, consisting of five years of compulsory elementary education (grade 1 to 5), four years of compulsory lower secondary education (grade 6 to 9) and two years of

non-compulsory upper secondary education (grade 10 to 11).³ The regular age of entry to the first grade is six years. Therefore, if children are not retained and do not interrupt their school career, they complete lower secondary education at ages 14-15 and upper secondary education when they are 16-17 years old.

The educational system in Colombia is a comprehensive school system with no tracking at any grade.⁴ Completion of the lower secondary cycle leads to a Certificate of Basic Baccalaureate Studies (*Certificado de Estudios de Bachillerato Básico*). Upon completion of the grade 11 of secondary school, all students must pass a national standardised exam (*SABER11*) to be awarded the title of Baccalaureate (*Título de Bachiller*), which gives access to higher education. This credential is equivalent to a US high-school diploma.

Although constitutional regulations in Colombia establish that elementary and lower secondary education are free and compulsory, the capacity of the system is in practice insufficient to accommodate full enrolment. In addition, dropout and inadequate students' progress in lower secondary education remains a bottleneck for enrolment. Net enrolment rates at all levels of secondary education increased from 64.1 percent in 2004 to 79.7 percent in 2013. However, enrolment levels vary widely across the country, particularly between rural and urban areas, as shown in Figure 3.1 (OECD, 2016).

Figure 3.1: Students who remain enrolled in education



Notes: Source: OECD, Education in Colombia Highlights 2016. Data correspond to year 2008.

³ Elementary and secondary education in Colombia are offered in two school calendars: Calendar “A” that runs from February until November, and calendar “B” from September to June. Most schools (92%) in the country operate in calendar A. Formal education is also offered by schools in three different class-schedules: a morning schedule, an afternoon schedule, and a full-day schedule. Students opt or are allocated by the school to attend either one of these. Most students in secondary education attend school either in the morning or the afternoon schedule (78%).

⁴ The Ministry of Education regulates all levels of education for both public and private schools.

3.3.2 The Free Retention Policy (FRP) reform

In 2002, by mandate of the Ministry of Education (Decree 230 of 2002), schools were each year restricted to retain up to a maximum of 5 percent of their students. This retention policy was implemented to reduce costs attributed to higher retention rates (i.e. low performance, low motivation, etcetera) without compromising the quality of education provided by the system (Martínez and Herrera, 2002). According to the policy mandate, a student should be retained if at least one of the following three circumstances holds: i) the student received an unsatisfactory performance evaluation in three or more school subjects in the current school year, ii) the student received an unsatisfactory performance evaluation in math and/or language courses during the current and two previous grades, or iii) the student failed to attend at least 25 percent of all school activities during the current school year. However, schools were required to adjust their evaluation standards to comply with the law, which forced them to promote at least 95 percent of all their students each year.

While the 5-percent retention rule was considered as moderately successful in increasing school enrolment, the incentives to underperform at school, as perceived by schools and teachers, led the Ministry of Education to revoke this retention restriction.⁵ In February 2009, the 5-percent retention restriction was ended by the Ministry of Education through a new regulation mandate (Decree 1290 of 2009), allowing schools from 2010 onwards to retain as many students as they considered necessary, and thereby giving them more discretion in their evaluation and promotion strategies. We use the term *Free Retention Policy* (FRP) reform to refer to this policy change. Overall, this reform increased students' retention rates across all grades of secondary education from 4.3 percent to 7.7 percent, on average, in all schools in the country.

3.4 Empirical Strategy

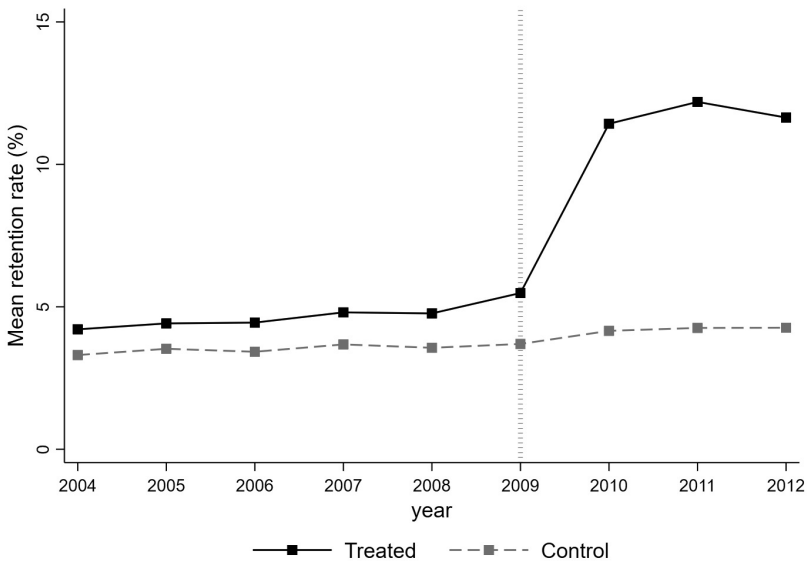
To estimate the effects of higher grade retention on school dropout rates, we use a difference-in-differences approach that exploits variation in retention rates at the school level, as induced by the FRP reform in Colombia. Since similar schools reacted differently to the policy change, our approach compares the dropout rates by grade in secondary school before and after the FRP reform, between schools where retention rates increased a lot (i.e. the treatment group) and schools where retention rates remained relatively constant across both periods (i.e. the control group).

Following the same strategy to identify treatment and control schools as in Chapter 2, we estimate difference-in-differences models that first employ a dichotomous treatment group variable, and then a categorical heterogeneous treatment variable. To generate the dichotomous treatment variable, we rank all schools by the percentage-point change in the retention rates between the periods before (2004-2009) and after the FRP reform (2010-2012). We define schools in which the increase in retention was above the median as the treatment

⁵ Ministry of Education, Press Release April 17, 2009.

group, since these schools were most relieved by the abolishment of the retention restriction. We define schools in which the increase in retention was below the median as the control group.⁶ Figure 3.2 shows the average retention rates for treated and control schools across years. First, we observe that between 2004 and 2009 in the treatment group, retention rates were about 1 percentage point higher than in the control group, which suggests that the schools most relieved by the treatment had retention rates nearer to the 5 percent ceiling. Second, we see that in control schools, on average, overall retention rates increased after the FRP reform by only 0.49 percentage points. In contrast, treated schools increased retention rates by 7.1 percentage points, i.e. 6.61 percentage points higher than at control schools.⁷ This pattern also holds for each grade from grade 6 to 11, as presented in Figure 3.3. Nonetheless, as the figure shows, earlier grades experienced greater growth in retention rates after the FRP reform, while in grade 11 the increase in retention was rather small (1.5 pp).

Figure 3.2: Retention rates by treatment status

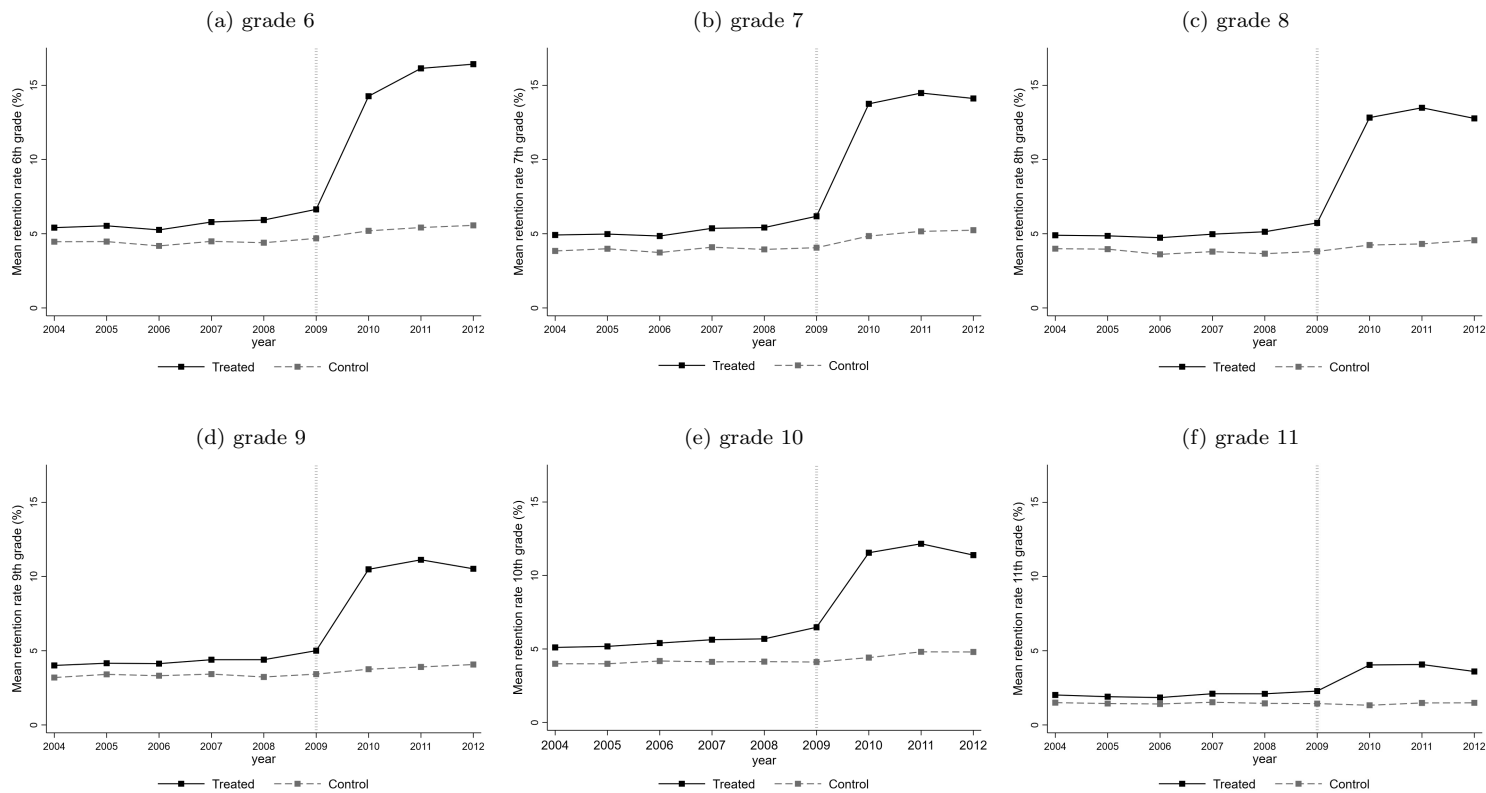


Notes: This figure shows school average retention rates in year t for treated and control schools. Averages include all students in grades 6 to 11.

⁶ By choosing this approach, we follow Havnes and Mogstad (2011) and Bauernschuster et al. (2016), who use a similar identification strategy for analyzing the effects of universal child care on parental economic outcomes.

⁷ Figure 3.A1 in the Appendix shows that retention rates were always higher among male students than female students, and that the difference in retention between treated and control schools before and after the FRP reform was driven more strongly by male students (4.02 pp) than female students (2.59 pp).

Figure 3.3: Retention rates by grade



Notes: This figure shows school average retention rates in year t including all students in each grade by treated and control schools.

To generate a categorical heterogeneous treatment variable, we followed the same procedure as in Chapter 2. Based on the increase in retention rates as explained above, we identify four groups: one control and three more and less intensely treated groups – high, mid and low treated schools. Figure 3.A2 in the Appendix shows the retention rates across time for these heterogeneously treated groups. We observe that retention rates among control schools barely changed: there was a slight decrease of 0.23 pp after the FRP reform. During the same period, low-treated, mid-treated and high-treated schools experienced increases in their grade retention rates of 1.82 pp, 4.29 pp and 9.72 pp respectively.⁸

We will distinguish between two outcomes as the effects of the reform. First, we consider *end-of-year dropout rates* (YD_{sgt}), referring to the proportion of pupils from a cohort enrolled in a given grade, who concluded the school year but were retained, and did not register in the educational system to repeat the failed grade. Since this is the rate of retained students who stop their school education in the same year of retention, we measure end-of-year dropout rates in year t as the outcome of retention rates during the same year t . Second, we consider *early dropout rates* (ED_{sgt}) that refer to the proportion of students in each grade who abandoned the educational system without concluding the school year. This is the dropout rate among all students enrolled in year $t + 1$, including those non-retained the preceding year and those repeating the failed grade. This output gives some indication of the spill-over effects of retention on non-retained students. Therefore, early dropout rates in year t are measured as the outcome of retention rates in year $t - 1$. Because we cannot identify students who permanently leave education and those who just interrupt their studies for at least a year and later return to school, both outcomes can only be interpreted as short-term (one year) effects of retention.

Our basic difference-in-differences models can then be expressed as:

$$YD_{sgt} = \alpha_{sg} + \delta_t + \gamma_{sg} [\alpha_{sg} \times LT] + \theta_{1st} [Group_s \times FRP_t] + \beta X_{sgt} + \varepsilon_{sgt} \quad (3.1)$$

$$ED_{sgt} = \alpha_{sg} + \delta_t + \gamma_{sg} [\alpha_{sg} \times LT] + \theta_{2st} [Group_s \times FRP_{t-1}] + \beta X_{sgt} + \varepsilon_{sgt} \quad (3.2)$$

where YD_{sgt} is the end-of-year dropout rate and ED_{sgt} the early dropout rate, as explained in the previous paragraph. s indexes school, g indexes grade (from 6 to 11), t indexes year. FRP is a dummy variable equal to 1 if the observation is from the period after the FRP reform (and 0 if the observation is from the period before the FRP reform). $Group$ is the treatment variable that takes value 1 for schools in the treated group, and zero for schools in the control group in our basic binary treatment specification; and a categorical variable that takes values from 1 to 4 corresponding to control, low-treated, mid-treated, and high-treated schools respectively. We estimate the model with and without a set of covariates for school, grade, teachers, and students' characteristics X , as described later in Table 3.1. All our estimations include year-fixed effects (δ), school-grade fixed

⁸ We observe a similar pattern for male and female students, although increases in retention are more pronounced among male students. While female retention rates changed after the FRP reform by -0.11 pp, 0.68 pp, 1.71 pp, and 4.07 pp for control, low-treated, mid-treated and high-treated schools respectively; male retention rates changed by -0.12 pp, 1.14 pp, 2.59 pp, and 5.65 pp for the corresponding treatment groups.

effects (α) and school-grade specific linear trends (γ). All standard errors are clustered at the school-grade level to ensure that we account for potential serial correlation, as indicated in the difference-in-differences literature (Bertrand et al., 2004). The models are weighted by the total number of students per year and school-grade.

Because of our treatment-control schools classification, θ_{st} are best interpreted as intention-to-treat effects (ITT). To obtain the average treatment effect on the treated (ATT) interpretation, we rescale the resulting coefficients by the difference in the retention rates between treated and control schools before and after the FRP reform. It is also important to note that our approach defines control and treatment groups (and more and less intensely treated groups) based on variation in retention rates by school over the sample period. Thus, the validity of our empirical design depends on the variation in retention rates being orthogonal to expected changes in dropout rates. We will investigate whether this key identifying assumption holds, estimating the following two specifications, taking advantage of the several years of data we have for the period before the reform:

$$YD_{sgt} = \alpha_{sg} + \delta_t + \gamma_{sg} [\alpha_{sg} \times LT] + \sum_{t=2005}^{2012} \theta_{4t} [Group_s \times \delta_t] + \beta X_{sgt} + \varepsilon_{sgt} \quad (3.3)$$

$$ED_{sgt} = \alpha_{sg} + \delta_t + \gamma_{sg} [\alpha_{sg} \times LT] + \sum_{t=2004}^{2013} \theta_{3t} [Group_s \times \delta_t] + \beta X_{sgt} + \varepsilon_{sgt} \quad (3.4)$$

With equations (3.3) and (3.4), we basically test for the assumption that treatment and control schools shared a common trend in end-of-year dropout and early dropout rates before the FRP reform, and that these trends would have remained the same if the retention policy would not have been changed. Therefore, the null hypothesis of interest here is that pre-treatment differences in trends (θ_t) between treated and control schools are not significantly different from zero.⁹ As we will elaborate later, we cannot reject such hypotheses at conventional significance levels in the case of early dropout rates (ED_{sgt}); however, we find that it is due to a significant and robust anticipatory effect in year 2010 that needs to be accounted for in the estimation of our main results.

In addition to specifications (3.3) and (3.4), we also perform several placebo tests, in which we assume that the FRP reform was introduced in different years before or after it was implemented. We examine whether such placebo reforms have any effects on our outcomes of interest. By not being able to reject the null hypothesis that such effects are significantly equal to zero in each of the estimated placebo cases, we can confirm that the estimated effects on end-of-year and early dropout rates can be attributed to the FRP reform only.

3.5 Data and Descriptive Statistics

3.5.1 Data

This study is based on information from 2004 to 2013 contained in the schools' national census, an administrative database that the Colombian Ministry of Ed-

⁹ That is, $\theta_{2005} = \theta_{2006} = \theta_{2007} = \theta_{2008} = \theta_{2009} = 0$, θ_{2004} being the base year of comparison in equation (3.3) for YD_{sgt} and $\theta_{2006} = \theta_{2007} = \theta_{2008} = \theta_{2009} = \theta_{2010} = 0$, θ_{2005} being the base year of comparison in equation (3.4) for ED_{sgt} .

ucation releases for public use through the National Administrative Department of Statistics (*DANE*). This census contains several school and organisational data that all schools in the country are compelled to report at the grade level on a yearly basis. We use information on the number of students by grade that are retained by the end of the school year, and those who abandon school before the school year ends (early dropouts). In addition, we are able to compute the number of retained students in year t that did not enrol in the educational system in year $t + 1$, which will give us insight in the percentage of retained students that do not continue with their school studies (end-of-year dropout rates). For retention and early dropout rates we can distinguish between male and female students; however, because of data limitations, we cannot analyse gender differences for end-of-year dropout rates. We use additional information on several student, teacher and school characteristics as covariates.

Figure 3.4: Retention and dropout rates by grade

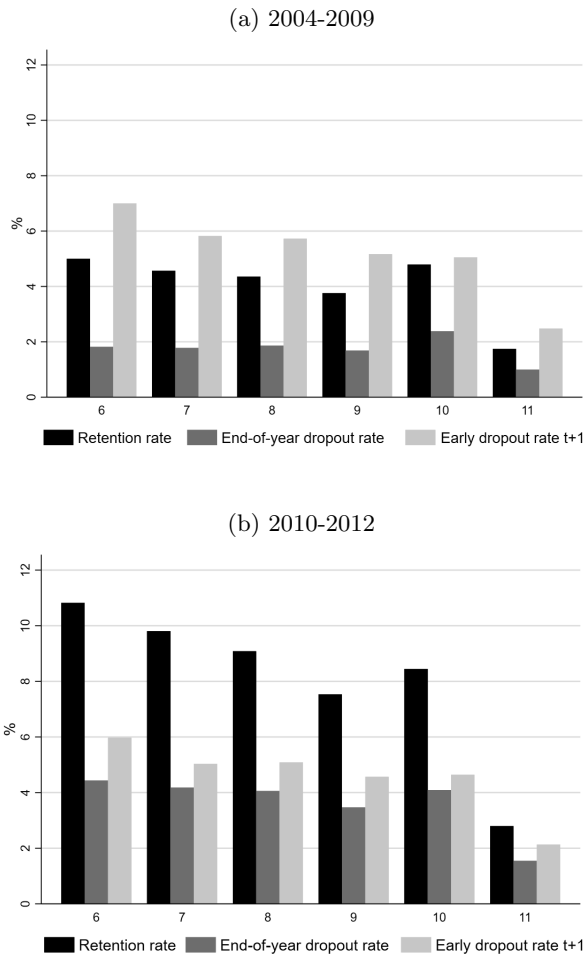
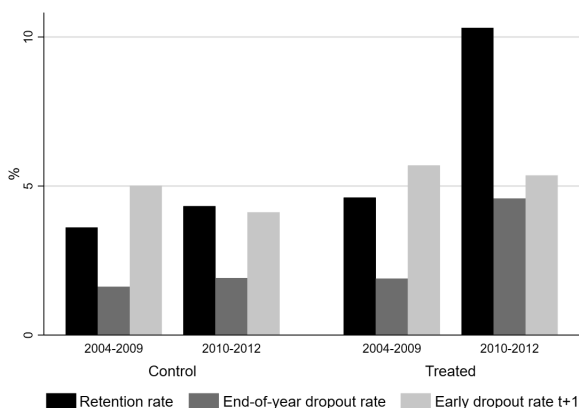


Figure 3.5: Retention and dropout rates by treatment status



Our unit of analysis is a school-grade-year combination for all schools that offer secondary education in the country. The estimation sample includes schools i) that offer education exclusively in Calendar A (February to November), ii) did not change this calendar during the period 2004-2013, iii) had no missing values on retention rates, dropout rates, and school covariates, and iv) reported information on retention rates at least one year before and one year after the FRP reform. The resulting dataset consists of an unbalanced panel of 9,449 schools, which in total contains 52,037 school-grade combinations and 368,433 schools-grade-year observations across the 2004-2013 period.

The major advantage of these administrative records is the ability to track the number of students by grade who exit from the universe of enrolled students from one year to the next. The second advantage is that the transfer of students between schools is accounted separately; hence, our dropout rates are purely the share of students who do not continue with their education. However, as mentioned in our empirical strategy, we can only identify the short-term (one-year) effects of retention on dropouts since we cannot distinguish between those who left school permanently or just temporarily.

As we can observe in Panel (a) of Figure 3.4, before the FRP reform, the average retention rates of students in grade 6 to grade 11 were kept at or below 5 percent, according to the law restriction. Panel (b) of the figure shows that retention rates nearly doubled for all grades from grade 6 to 10 after the reform. Only at grade 11, the last level of secondary education, retention rates were below 2 percent before the reform, increasing by just 1 percentage point after the reform.

Figure 3.4 also shows that end-of-year dropout rates by grade were slightly less than half of the retention rates before the reform. This means that, between 2004 and 2009, less than fifty percent of the retained students in each grade dropped out of school by the end of the year of their retention. With the increase of retention rates after the reform, all end-of-year dropout rates in the period 2010-2012 more than doubled, with the only exception of grade 11.

Finally, regarding early dropout rates in year $t + 1$, Figure 3.2 reveals that even before the reform those rates were high and close to 6 percent. These dropout rates, however, decreased after the FRP reform by nearly 1 percentage point in each grade.

More interestingly, when we compare the average retention and dropout rates between treated and control schools, we find first suggestive evidence of a positive effect of increased retention on end-of-year dropout rates. As figure 3.5 shows, whereas end-of-year dropout rates in control schools remained almost the same before and after the reform, these rates grew by approximately 3.5 percentage points in treated schools between the same two periods. The descriptive evidence on early dropout rates is, however, not that clear. Although early dropout rates among schools in the treatment group decreased less than they did in control schools between the periods before and after the reform, the difference seems to be very small. Estimation results in Section 3.6 will reveal whether this difference holds in our difference-in-differences setting.

3.5.2 Descriptive Statistics

Table 3.1 shows pre-treatment summary statistics for our outcome variables, retention rates, and the set of covariates included in our models for both the treatment and control group of schools. First, we observe that between 2004 and 2009 in the treatment group, retention rates were about 1 percentage point higher than in the control group, suggesting that the schools most relieved by the treatment had retention rates closer to the 5 percent limit set by the law. As expected, the average percentage of repeaters in each grade in year $t + 1$ was higher in the treatment group as well. Second, we observe that both end-of-year dropout rates and early dropout rates were also slightly higher among schools in the treatment group prior to the reform.

Our difference-in-differences approach identifies the effects of retention by comparing the change in dropout rates before and after the FRP reform among treatment and control schools. Table 3.1 shows that the treatment and control schools have fairly similar characteristics. Interestingly, there are no noticeable differences in the share of female students in the grades of the treatment and control-group schools, nor is there significant disparity in the share of subsidised students, or the share of students from ethnic minorities, with disabilities or exceptional abilities, or the share of students documented as victims of the armed conflict. Moreover, the difference in the share of rural and urban schools, the percentage of schools that apply non-traditional teaching methods, the share of teachers with tertiary and pedagogical education as well as the number of students per medical and administrative personnel do not significantly differ between treated and control schools. Further, there are no substantial differences in the percentage of teachers hired under the new pay-scale regulation introduced in Colombia in 2002.¹⁰

¹⁰ In 2002, by mandate of the Ministry of Education (Decree 1278 de 2002) the career and pay scale of public-school teachers was reformed through the introduction of a selective entry test and further quality incentives. The regulation applied only to newly hired teachers, creating a mix of new-pay-scale regulation and old-pay-scale regulation teachers in Colombian schools.

Table 3.1: Pre-treatment summary statistics

	Control Mean	Treated Mean	All Min	All Max	Difference (C) - (T) b	se
Retention rate	3.66	4.64	0	100	-0.98***	0.03
Male retention rate	2.28	2.94	0	100	-0.66***	0.02
Female retention rate	1.38	1.70	0	100	-0.32***	0.01
End-of-year dropout rate t	1.62	1.91	-298	100	-0.29***	0.03
Share of repeaters $t+1$	2.03	2.64	0	100	-0.61***	0.03
Early dropout rate $t+1$	4.82	5.66	0	100	-0.85***	0.04
Male early dropout rate $t+1$	2.61	3.09	0	100	-0.48***	0.02
Female early dropout rate $t+1$	2.21	2.57	0	100	-0.37***	0.02
<i>Covariates</i>						
Average age	14.07	14.19	9	20	-0.12**	0.05
Share of female students	50.94	50.61	0	100	0.34	0.22
Share of subsidised students	4.61	4.43	0	100	0.18	0.21
Share of students from ethnic minorities	5.10	5.13	0	100	-0.03	0.25
Share of students victims of armed conflict	1.32	1.44	0	100	-0.12	0.18
Share of students with disabilities	1.02	1.01	0	100	0.01	0.02
Share of students with exceptional abilities	0.85	0.84	0	73	0.01	0.01
Rural school	0.34	0.33	0	1	-0.01	0.02
Private school	0.28	0.22	0	1	0.06***	0.01
Non-traditional teaching school	0.11	0.09	0	1	0.02	0.02
Number of class schedules	1.58	1.74	1	5	-0.16**	0.06
Number of groups per grade	1.98	2.49	1	45	-0.51**	0.19
Average class size	28.89	30.15	1	202	-1.26	0.94
Total students per grade	57.20	75.26	1	1,204	-18.06***	1.03
Total students per school	412.14	524.90	4	6,330	-112.75***	5.95
Share of teachers with pedagogical education	86.19	87.42	0	100	-1.23	0.81
Share of teachers with tertiary education	94.79	95.88	0	100	-1.09	0.63
Share of teachers under new pay-scale	37.97	40.31	0	100	-2.34*	1.09
Students per teacher	20.33	21.70	1	263	-1.37**	0.51
Students per administrative personnel	77.55	80.98	1	5,166	-3.43	2.46
Students per medical personnel	197.19	208.02	1	5,166	-10.83	6.96
Observations	90,248	93,478	183,726			
School-grades	24,970	25,985	50,952			
Schools	4,668	4,741	9,409			

There are, however, some notable significant differences between treatment and comparison schools. First, although the difference in average age of students and number of class schedules per school are statistically significant with 95 percent of confidence, the size of these differences is that small that they are not economically significant: students in treatment and control schools are on average 14 years old, and schools in treatment and control groups offer education in either one or two class schedules. Most importantly, in the treatment group the share of private schools is 6 percentage points lower than in the control group. This is not surprising since the expansion of retention rates is likely to be stronger in public schools where retention rates before the reform were more restricted and strictly controlled by the government as public funds are conditioned to the annual information reported by schools. In addition, it might be related to the higher socio-economic status of students in private schools, which could make retention rates less likely to increase after the reform.

In relation to the proportion of public and private schools in treatment and control groups, we also observe significant differences in the total number of students per grade and school as well as in the number of groups per grade. This might be due to the discreteness to increase retention after the reform; retention rates are more likely to increase in larger than smaller schools, and in Colombia public schools tend to be larger, on average, than private schools. Since the characteristic of private or public school does not change over time in our sample, these significant differences will most probably be picked up by the school-grade fixed effects and school-grade specific linear trends. It is also important to note in this respect that the average number of students per class does not differ between treated and control schools, despite the significant differences previously mentioned. Finally, to further investigate the robustness of our results with respect to time-varying school characteristics, we will run the regressions both without any covariates and with a rich set of school-grade specific time-varying covariates. As will be shown in Section 6, the inclusion of covariates does not affect the positive sign or the size of our estimates.

3.5.3 Common Trends: Graphical evidence

The key identifying assumption of any difference-in-differences model is the common trend assumption. In our case this means that, conditional on school-grade fixed effects and the set of time-varying covariates, there are no unobserved characteristics of a school-grade that vary over time and are correlated with school retention rate increases and future changes in the schools' dropout rates.

As we showed in the preceding section, the treatment and control schools are very similar in their pre-reform characteristics. Despite this, it might be the case that time-varying factors that are correlated with dropout rates evolve differently in treated schools as compared to control schools and thus bias the estimates. To address this concern, we investigate whether treatment and control schools show the same trend in both our outcomes of interest during the pre-treatment period. In this section, we perform a first graphic check of the common trend and in Section 3.6 we will show the results of the regressions tests defined in equations (3.3) and (3.4).

Figure 3.6 shows the average dropout rates across time for treated and control schools. First, the graph in panel (a) provides suggestive evidence of a common trend between treated and control schools during the pre-treatment period regarding end-of-year dropout rates. There seems to be a small anticipatory effect one year prior to the treatment: if students that concluded the school year but were retained at the end of 2009 knew that the next year the probabilities of retention were going to be even higher, those at the lower end of the ability distribution or those who were less confident about their academic success they might have felt discouraged to repeat the grade and dropped out of school. The graph in Panel (a) of Figure 3.6 also suggests a substantial positive effect of retention on end-of-year dropout rates among the retained students. Figure 3.A3 in the Appendix indicates that the same conclusion holds if we use the heterogeneous treatment groups low-treated, mid-treated, and high-treated. The effect of higher retention on end-of-year dropout rates seems to increase proportionally to the intensity of the treatment.

Second, Panel (b) of Figure 3.6 suggests the existence of a common trend for early dropout rates during the pre-reform period (2005-2010), except for year 2010. This exception is likely to be an anticipatory effect of the FRP reform: since enrolled students were aware since the beginning of the year that the probabilities of being retained were going to be much higher than in the past, those at the lower end of the ability distribution or those who were less confident about their school success might have made the decision of dropping out of school before the termination of the school year. Panel (b) of Figure 3.6 also suggests a positive effect of retention on early dropout rates; the difference in early dropout rates between treated and control schools in the pre-reform period increases consistently after 2010.¹¹ As shown in Figure 3.A5 in the Appendix, the same inferences hold when using the heterogeneous treatment variable instead of the binary variable. The effect of higher retention on early dropouts seems to be linear; there is no apparent effect among low-treated schools whereas the effect is positive and increasing among mid-treated and high-treated schools respectively.

Graphically, we can also observe some differences between grades. Figure 3.7 shows the graphical tests of the common trend for end-of-year dropout rates by grade. This figure suggests once more the existence of a common trend at each grade between control and treated schools for the period 2004-2009. The small anticipatory effect appears here as well, except in grade 6 where no anticipatory effect seemed to have occurred. Figure 3.7 is preliminary descriptive evidence of a substantial positive effect of higher retention on end-of-year dropout rates among retained students in all grades, even in grades 10 and 11. This effect, however, appears to be stronger at earlier grades, and smallest at grade 11.

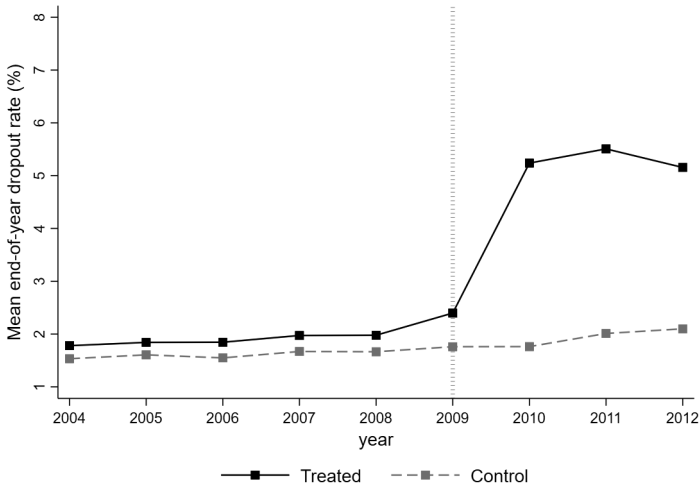
Finally, Figure 3.8 shows the average early dropout rates across time by grade. In general, the figure indicates that common trends hold in each grade

¹¹ Figure 3.A4 in the Appendix suggests that the common trend in early dropout rates between treated and control schools also holds among female and male students, separately. Even though male early dropout rates were slightly higher than female dropout rates during the entire period, the figure shows the potential positive effect of retention on early dropout rates to be similar among female and male students: the difference between control and treated groups nearly doubled after the FRP reform. The figure also suggests that the anticipatory effect of the reform in 2010 was more pronounced among male than female students.

between control and treated schools during the pre-treatment period. The anticipatory effect only seemed to have occurred from grade 6 to grade 9. The figure also indicates that early dropout rates are larger at earlier grades and that the grade 11 is the grade with the lowest percentage of dropouts across time. The effect of retention on dropout rates seems to be positive and strongest at the first level of secondary education (grade 6) whereas there does not seem to be any effect at grades 10 and 11.

Figure 3.6: Dropout rates by treatment status

(a) End-of-year dropout rates t



(b) Early dropout rates $t+1$

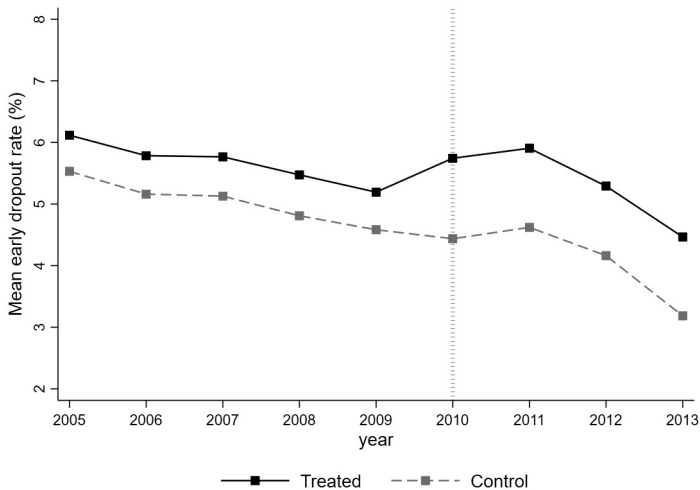
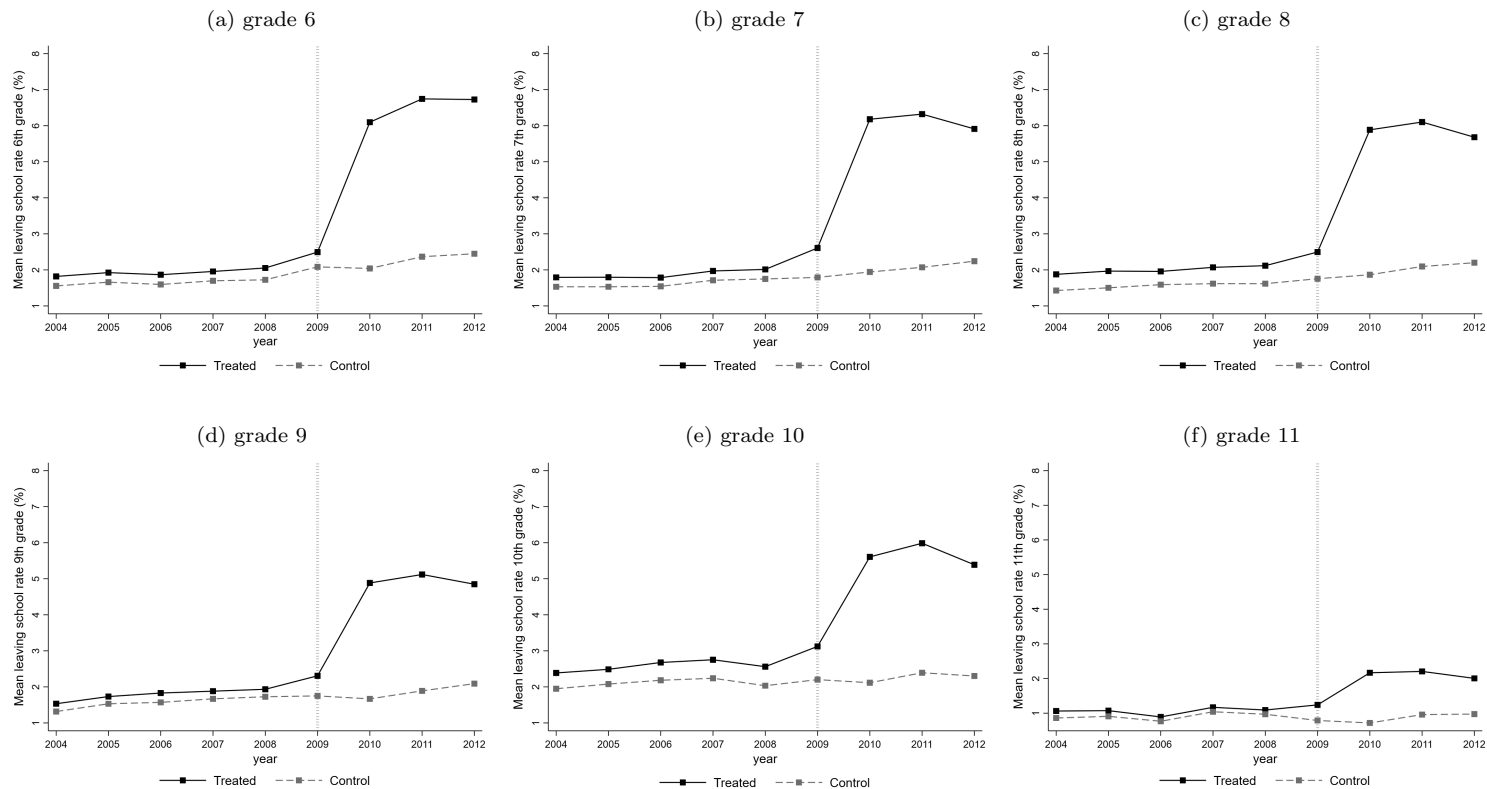
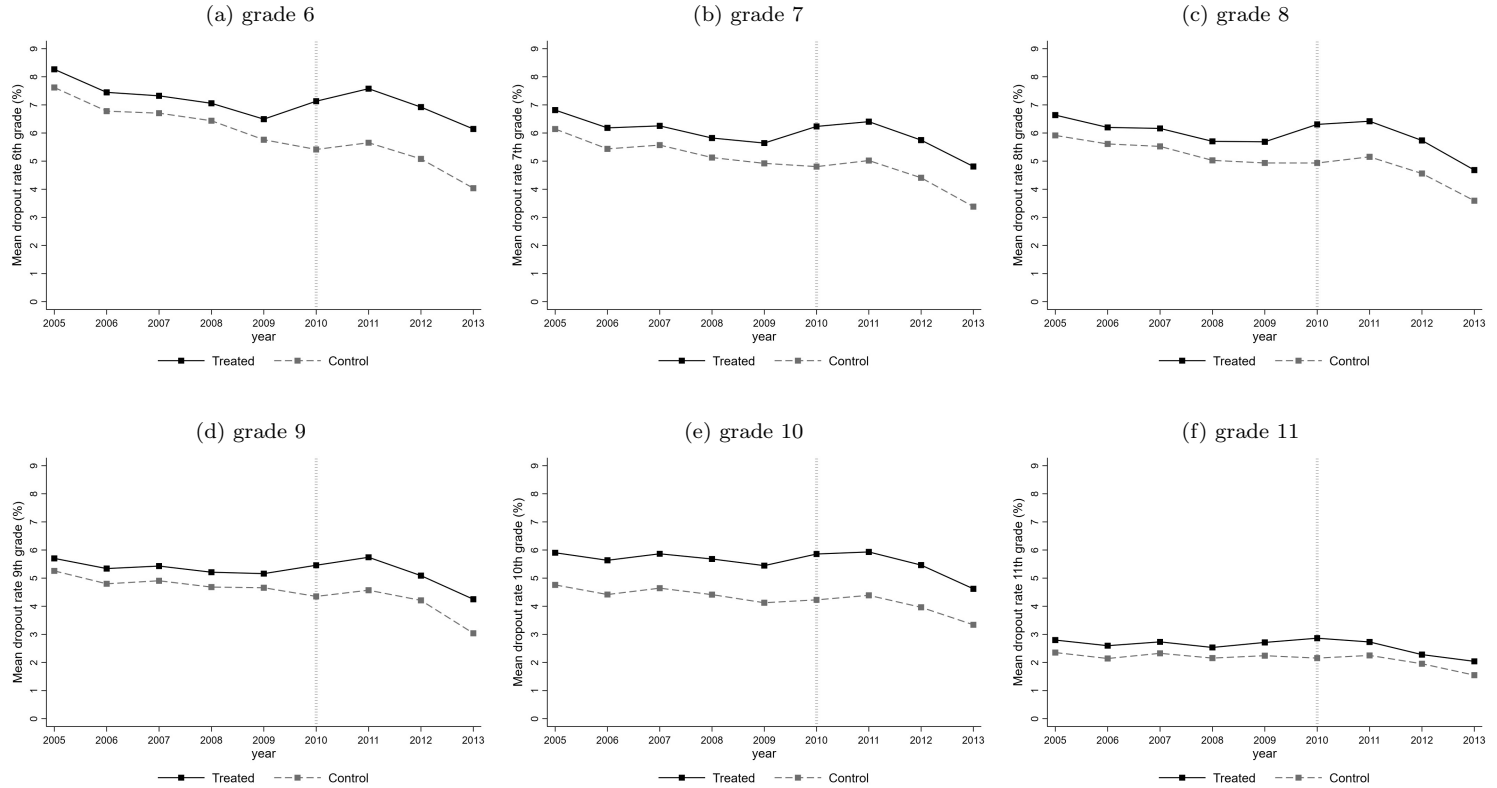


Figure 3.7: End-of-year dropout rates by grade t



Notes: This figure shows school average end-of-year dropout rates in year t including all students in each grade by treated and control schools.

Figure 3.8: Early dropout rates by grade $t+1$



Notes: This figure shows school average dropout rates in year $t+1$ including all students in each grade by treated and control schools.

3.6 Results

3.6.1 End-of-year dropout rates: Testing the common trend assumption and anticipatory effect

A precondition for the validity of the difference-in-differences estimator is that the treatment is not implemented based on pre-existing differences in outcomes between treatment and control groups. The graphical evidence presented in the previous section suggests the existence of such common trend in early dropout rates between treated and control schools before the implementation of the FRP. However, the graphs also suggest the existence of an anticipatory effect, which could affect our main estimations, given the fact that the reform was announced one year before it came into force.

In this section, we present the results of the difference-in-differences common trend test for end-of-year dropout rates. This refers to the proportion of pupils from a cohort enrolled in a given grade, who concluded the school year but were retained, and did not register in the educational system to repeat the failed grade. As explained in Section 3.4, since this is the rate of students who stop their school education in the same year of retention, we measure end-of-year dropout rates in year t as the outcome of retention rates during the same year t .¹²

The graphical evidence presented in Section 3.5.3 suggested the existence of a common trend in end-of-year dropout rates between treated and control schools before the implementation of the FRP; however, the graphs also suggest the existence of a tiny anticipatory effect, which could affect our main estimations, given the fact that the reform was announced one year before it came into force.

In Table 3.2 we show the common trend estimates for end-of-year dropout rates before and after the FRP reform, as described in equation (3.3). As required, the pre-treatment trends from 2005 to 2008 are not significantly different from the baseline year 2004; the corresponding coefficients are in fact close to zero. On the contrary, the post-treatment trends from 2010 to 2012 are all statistically significant and positive. The size of these estimates suggests an important positive impact of the FRP reform of approximately 3 percentage points higher end-of-year dropout rates among the retained students.

However, it is also important to note that the coefficient for the pre-treatment trend in year 2009 differs with 90 percent of confidence from the baseline. Although the size of this coefficient is very small in comparison with the subsequent post-treatment trend estimates, it suggests the possibility of an anticipatory effect in end-of-year dropout rates in year 2009. When using the heterogeneous treatment variable instead of the binary indicator, as shown in Table 3.A1 in the Appendix, we observe that the marginal anticipatory effect in end-of-year dropout rates is entirely driven by schools in the high-treated group. In addition, these results in Table 3.A1 show that the pre-treatment trends are also similar between the control and different-intensity treated schools. Finally, the FRP

¹² Because of data limitations, as mentioned earlier in Section 3.5.1, for this outcome of end-of-year dropout rates we cannot distinguish between male and female students; therefore, gender analyses are not possible.

trends estimates in this table suggest that the effect of the reform on end-of-year dropouts is linear to the intensity of treatment.

Table 3.2: Common trend test - End-of-year dropout rate ϵ_t

	(1)	(2)
<i>Pre-treatment trends</i>		
Treated \times 2005	-0.091 (0.114)	-0.155 (0.108)
Treated \times 2006	0.053 (0.110)	-0.009 (0.103)
Treated \times 2007	0.059 (0.117)	-0.018 (0.111)
Treated \times 2008	0.088 (0.117)	0.016 (0.113)
Treated \times 2009	0.310** (0.124)	0.231* (0.121)
<i>FRP trends</i>		
Treated \times 2010	3.163*** (0.147)	3.065*** (0.144)
Treated \times 2011	3.173*** (0.145)	3.061*** (0.139)
Treated \times 2012	2.790*** (0.152)	2.720*** (0.146)
Covariates	No	Yes
Schools	9,782	9,449
School-grades	53,867	52,037
Observations	389,775	368,433
Adjusted R^2	0.171	0.183

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To further check for the potential existence of the anticipatory effect and the impact it could have on our main estimates, we run some additional placebo tests. We first assume artificially that the FRP reform was introduced in different years before it actually occurred and estimate equation (3.1) for the period 2004-2009. Table 3.3 shows the resulting estimates of this exercise in columns (1) and (2). We observe that the last two placebo coefficients are positive and significant, which means we cannot confirm the existence of a common trend in years 2008 and 2009.

Since we suspect an anticipatory effect in end-of-year dropout rates occurring in 2009, based on the previous common trend tests, in columns (3) and (4) of Table 3.3 we present the results of the same placebo tests excluding year 2009 from the sample. The results show again that excluding the year of the anticipatory effect leads to accept the null hypothesis of a common trend in all the necessary cases; the estimates from all placebo tests are then statistically zero, as required to be able to use a difference-in-differences strategy. Finally, in columns (5) and (6) of Table 3.3, we present the outcomes of the two possible placebo tests for the after-reform period (2010-2012), which also give the non-significant required estimates to be able to use our difference-in-differences approach. Given the results of this section, we consider the anticipatory effect in end-of-year dropout rates to be likely. We will therefore estimate our main

difference-in-differences model (1) both including and excluding data from year 2009 and show the results in the next Section.

Table 3.3: Placebo tests - End-of-year dropout rate t

	Including year 2009		Excluding year 2009		Year > 2009	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated \times FakeFRP ₂₀₀₅	0.009 (0.062)	0.006 (0.067)	-0.026 (0.064)	-0.013 (0.069)		
Treated \times FakeFRP ₂₀₀₆	0.071 (0.053)	0.038 (0.068)	0.028 (0.056)	-0.026 (0.070)		
Treated \times FakeFRP ₂₀₀₇	0.030 (0.053)	0.029 (0.068)	-0.040 (0.057)	-0.061 (0.063)		
Treated \times FakeFRP ₂₀₀₈	0.126** (0.052)	0.138** (0.056)	0.053 (0.063)	0.031 (0.066)		
Treated \times FakeFRP ₂₀₀₉	0.187*** (0.063)	0.189*** (0.065)				
Treated \times FakeFRP ₂₀₁₁					0.322 (0.251)	0.230 (0.250)
Treated \times FakeFRP ₂₀₁₂					-0.322 (0.250)	-0.230 (0.251)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	52,674	50,109	51,356	48,995	48,063	46,179
Observations	233,819	201,385	195,105	179,687	131,440	126,563

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6.2 End-of-year dropout rates: The effect of the FRP reform and increased retention

Table 3.4 presents the baseline estimates of the FRP reform on end-of-year dropout rates among retained students at secondary school. As indicated in Section 3.4, we present our main results both excluding and including time variant school-grade specific covariates. All our estimations include year-fixed effects, school-grade fixed effects and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade and standard errors are always clustered at the school-grade level to ensure that we account for potential serial correlation.

Table 3.4 provides evidence of a large and significant positive effect of the FRP reform on the average end-of-year dropout rate of secondary school retained students. On average, the rate of students that abandoned education after being retained at the end of the school year was 3 percentage points larger in treated than in control schools after the implementation of the FRP reform. Including or excluding year 2009 from the estimations, as shown in columns (3) and (4) of Table 3.4, does not significantly change the results.

Given the treatment classification and identification strategy we apply, our main estimates are best interpreted as intention-to-treat effects (ITT). To obtain the average treatment effects on the treated (ATT), we rescale these estimates

Table 3.4: Effect of the FRP reform on end-of-year dropout rate t

	Including year 2009		Excluding year 2009	
	(1)	(2)	(3)	(4)
Treated \times FRP	3.084*** (0.094)	3.084*** (0.106)	3.298*** (0.124)	3.427*** (0.149)
Covariates	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409
School-grades	52,674	50,109	52,674	50,109
Observations	359,457	313,053	316,845	272,017
Adjusted R^2	0.269	0.271	0.280	0.278

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

by the effect of the FRP reform on retention rates. Table 3.A2 in the Appendix presents the estimations of the latter effects. On average, after the FRP reform, retention in secondary schools increased by 7.2 percentage points. Based on the previous measures, we can conclude that the 7.2 percentage points increase in retention of secondary school students, as a product of the FRP reform, caused a rise in the dropout rate of retained students by the end of the school year of 3.1 percentage points. This means that approximately 40 percent of retained students after the reform decided not to continue with their secondary education, or at least interrupted their schooling temporarily, after learning they needed to repeat the grade.

To the best of our knowledge, there was not any other educational reform or regulation change around the same FRP period that could have affected dropout rates. Therefore, we discard the possibility that difference in end-of-year dropout rates between treated and control schools after the FRP reform are caused by any other reason than the increase in students' retention. This is a very relevant finding that points to the undesirable effect of retention as it seems to discourage retained students to continue with their school development. This finding is consistent with the literature (e.g Jacob and Lefgren, 2009; Manacorda, 2012) that has found that grade failure induces students to drop out at the end of the school year when failure occurs.

3.6.2.1 Heterogeneous effects by treatment intensity

To provide further insights into the effects of retention on early dropouts during secondary education, we examine whether our estimate of main interest differs across schools that were treated with different intensity. For this analysis, we use the heterogeneous treatment groups as explained in Section 3.4. We first estimate the effects of the FRP reform on retention across low-treated, mid-treated and high-treated schools (see Table 3.A3 in the Appendix), and then re-estimate equation (3.1) using this heterogeneous treatment variable instead of the binary indicator. The main results are presented in Table 3.5.

The results in Table 3.5 confirm our main finding of a large and significant positive significant positive effect of the FRP reform on end-of-year dropout

rates, proportional to the extent of retention growth in schools treated differently. On average, dropout rates of retained students by the end of the school year in high-treated, mid-treated and low-treated schools increased respectively by 4.7, 1.9 and 0.7 percentage points after the implementation of the FRP reform.

Rescaling these estimates to the effect of the FRP reform on retention increases,¹³ we can conclude that the effect of retention of end-of-year dropout rates among retained students is likely to be linear. This means that the largest dropout rates of retained students after the FRP reform was experienced by high-treated schools. In all treated schools, regardless of the retention rates increases after the FRP reform, on average, about 40 percent of retained students did not register to repeat the failed grade and did not continue with their education in the year following retention. This is at least a huge undesirable temporary short-term effect of retention. Further research on the long-term effects of retention on dropping out of school is essential to unveil the full dimension of impacts that retention may have on retained students.

Table 3.5: Effect of the FRP reform on end-of-year dropout rate ϵ_t

	Including year 2009		Excluding year 2009	
	(1)	(2)	(3)	(4)
Low-treated \times FRP	0.744*** (0.118)	0.671*** (0.135)	0.682*** (0.160)	0.623*** (0.190)
Mid-treated \times FRP	1.987*** (0.125)	1.920*** (0.141)	1.994*** (0.165)	1.992*** (0.198)
High-treated \times FRP	4.696*** (0.141)	4.664*** (0.160)	4.982*** (0.179)	5.158*** (0.218)
Covariates	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409
School-grades	52,674	50,109	52,674	50,109
Observations	359,457	313,053	316,845	272,017
Adjusted R^2	0.272	0.274	0.282	0.281

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6.2.2 Heterogeneous effects by grade

Our main estimates could also differ by grade if students experience retention differently depending on the timing of retention along the secondary school cycle. We may expect that retention at earlier grades affects students' attitudes and future school perspectives more strongly than retention at later stages, therefore, leading to larger effects on retained students' dropouts at the beginning of secondary education. To test for these potential differences, we re-estimate equation (3.1) for each grade separately, from grade 6 to grade 11 of secondary education, using both the binary and the heterogeneous treatment variables. The resulting estimates are shown in Table 3.6 and Table 3.7 respectively.

¹³ Table 3.A3 in the Appendix shows the FRP reform increased retention in low-treated, mid-treated and high-treated schools by 1.6, 4.4, and 11.1 percentage points respectively.

First, when using the binary treatment indicator as shown in Table 3.6, we find that the overall positive effect of the FRP reform on end-of-year dropout rates is significant across all six grades. We also observe that the effects of retention on retained dropouts is stronger and of similar magnitude among students from grade 6 to grade 8. This effect decreases slightly for retained students during grades 9 and 10 and it is smallest at grade 11, where retention increases after the FRP reform were also the smallest.

Nonetheless, when we rescale the estimates in Table 3.6 to the increases in retention due to the FRP reform (as shown in Table 3.A4 in the Appendix), we find that the effect is quite uniform for students in grades 6, 7, 8, and 10; but stronger if students are retained at the end of grade 9 and grade 11. After the implementation of the FRP reform, approximately 45 to 48 percent of students retained in grade 9 and grade 11 did not continue with their education after being retained. This finding indicates that retention has a tougher effect on students enrolled in the school year that would otherwise lead them to either (1) complete the lower secondary cycle and obtain the Certificate of Basic Baccalaureate Studies, that is, grade 9; or (2) complete and obtain the high-school diploma, that is, grade 11. This result suggests then that particular attention to retained students at these two specific grades is necessary since dropping out of school at these stages would not only be a high risk but would also imply the highest cost of retention due to the forgone opportunity for students to finalise either one of the two cycles of secondary education.

Finally, when using the low-, mid-, and high-treatment variables, as shown in Table 3.7, we observe large heterogeneity in the effect of retention across grades of schools treated with different intensity. Taking into account the increases in grade retention induced by the FRP reform (shown in Table 3.A4 in the Appendix), we can infer that at earlier grades (from grade 6 to grade 8), retention seems to affect end-of-year dropout rates linearly to the increases in retention: on average, after the implementation of the FRP reform, approximately 40 percent of students retained from grade 6 to grade 8 did not continue with their education after being retained.

However, retained students from grade 9 to grade 11 responded very differently in their decision of leaving school afterwards. These effects appear to be nonlinear, since retained students enrolled in grade 9 to grade 11 in mid-treated schools reacted with higher dropout rates (approximately 60 percent of retained students left school after retention) than similar students in high-treated schools where retention rates were even higher (approximately 48 percent of retained students left school after retention). This finding suggests that students retained at later stages of secondary education in a school environment where retention is less common can cause more harm to the future perspectives of those students than if they would have been in schools where retention was more pronounced or more likely to occur at the end of the school year. Further research into the effects of retention in different peer environment is therefore needed to better understand the decision of retained students about continuing or not with their schooling.

Table 3.6: Effect of the FRP reform on end-of-year dropout rate by grade t

	grade 6		grade 7		grade 8		grade 9		grade 10		grade 11	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated \times FRP	3.816*** (0.234)	3.725*** (0.269)	3.532*** (0.224)	3.458*** (0.246)	3.361*** (0.222)	3.517*** (0.252)	2.878*** (0.213)	2.917*** (0.238)	2.667*** (0.239)	2.669*** (0.267)	1.174*** (0.184)	1.229*** (0.215)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,731	9,353	9,664	9,292	9,531	9,174	9,327	8,976	7,839	7,576	7,628	7,388
Observations	67,683	57,977	66,246	57,271	64,180	55,846	61,663	53,859	51,014	45,053	48,671	43,047
Adjusted R^2	0.301	0.302	0.271	0.273	0.245	0.244	0.230	0.236	0.261	0.264	0.128	0.123

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.7: Effect of the FRP reform on end-of-year dropout rate by grade t - heterogeneous treatment

	grade 6		grade 7		grade 8		grade 9		grade 10		grade 11	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Low-treated \times FRP	1.091*** (0.168)	1.118*** (0.177)	0.923*** (0.160)	0.944*** (0.168)	0.955*** (0.156)	0.873*** (0.163)	1.064*** (0.154)	1.092*** (0.160)	1.026*** (0.177)	0.943*** (0.183)	0.481*** (0.112)	0.517*** (0.124)
Mid-treated \times FRP	2.766*** (0.186)	2.802*** (0.196)	2.442*** (0.173)	2.477*** (0.182)	2.350*** (0.165)	2.239*** (0.172)	2.330*** (0.159)	2.321*** (0.164)	2.269*** (0.192)	2.171*** (0.194)	0.656*** (0.122)	0.713*** (0.132)
High-treated \times FRP	5.720*** (0.223)	5.767*** (0.233)	5.306*** (0.212)	5.417*** (0.217)	5.041*** (0.209)	5.028*** (0.214)	4.347*** (0.200)	4.387*** (0.206)	4.552*** (0.218)	4.441*** (0.221)	2.056*** (0.158)	2.092*** (0.169)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,731	9,353	9,664	9,292	9,531	9,174	9,327	8,976	7,839	7,576	7,628	7,388
Observations	64,330	60,058	63,097	59,325	61,377	57,964	59,202	56,085	49,440	47,087	47,351	45,170
Adjusted R^2	0.249	0.257	0.237	0.245	0.220	0.227	0.206	0.211	0.230	0.240	0.114	0.120

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6.3 Early dropout rates: Testing the common trend assumption and anticipatory effects

In this section, we show a similar set of results than in Section 3.6.1 but for the second outcome of our interest, that is, early dropout rates in the year following retention.

Table 3.8 shows the common trend estimates for early dropout rates before and after the FRP reform. As required, the pre-treatment outcome trends from 2006 to 2009 are not significantly different from the baseline year 2005. In fact, these pre-treatment coefficients are very precise estimates of a zero difference in differences. Conversely, the trend estimates from 2011 to 2013 are all statistically significant and positive, although very small (lower than one percentage point) and decreasing over time. Overall, this result suggests a positive impact of the FRP reform and increased retention on early dropout rates.

However, Table 3.8 also shows that the early dropout rates in 2010, which are the result of retention in the pre-treatment year 2009, significantly differ between treated and control schools by almost one percentage point. These estimates are statistically significant and of the same size either excluding or including covariates in the model; and hold when distinguishing between male and female students as well.

As shown in Table 3.A6 in the Appendix, this anticipatory effect is also evident when using the heterogeneous treatment variable instead of the binary variable. The anticipatory effect in 2010 is mostly driven by schools in the mid-treated and high-treated groups; although, there is a significant anticipatory effect in early dropout rates among male students in low-treated schools. Table 3.A6 also confirms the validity of similar pre-treatment trends between the control and different-intensity treated schools. The effect of higher retention on early dropouts seems to be linear to the intensity of treatment; there is no apparent effect among low-treated schools whereas the effect is positive and increasing among mid-treated and high-treated schools respectively. These results suggest that the positive effect of retention on early dropout rates is strongest in high-treated schools and higher among male than female students.

Overall, we find a very robust indication of an anticipatory effect in early dropout rates one year prior to the reform implementation. As suggested by the common-trend tests, it is plausible that some students dropped out of school before concluding the school year as a reaction to the early announcement of the new law that would increase the likelihood of being retained in the following years.

To further check for the influence that such anticipatory effect could have on our main estimates, we run some additional placebo tests. First, we assume artificially that the FRP reform was introduced in different years before it actually occurred and estimate equation (3.2) for the period 2005-2010. The results are shown in Table 3.9. We observe that in all the five placebo tests performed, the estimated coefficients are positive and highly significant, regardless if we use the entire sample or the male and female students subsamples separately.

Table 3.8: Common trend tests - Early dropout rate $t+1$

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Pre-treatment trends</i>						
Treated \times 2006	0.029 (0.052)	0.029 (0.053)	0.024 (0.033)	0.028 (0.034)	0.005 (0.027)	0.001 (0.027)
Treated \times 2007	0.038 (0.053)	0.031 (0.054)	0.041 (0.036)	0.036 (0.035)	-0.003 (0.027)	-0.005 (0.027)
Treated \times 2008	0.065 (0.054)	-0.003 (0.055)	0.042 (0.034)	-0.005 (0.034)	0.023 (0.028)	0.003 (0.028)
Treated \times 2009	0.023 (0.053)	-0.003 (0.055)	0.031 (0.033)	0.021 (0.034)	-0.008 (0.027)	-0.024 (0.028)
Treated \times 2010	0.701*** (0.056)	0.652*** (0.058)	0.496*** (0.035)	0.451*** (0.036)	0.206*** (0.028)	0.201*** (0.029)
<i>FRP trends</i>						
Treated \times 2011	0.915*** (0.057)	0.853*** (0.058)	0.605*** (0.035)	0.554*** (0.035)	0.311*** (0.029)	0.299*** (0.030)
Treated \times 2012	0.430*** (0.047)	0.458*** (0.047)	0.260*** (0.029)	0.274*** (0.029)	0.170*** (0.025)	0.184*** (0.026)
Treated \times 2013	0.384*** (0.045)	0.404*** (0.045)	0.229*** (0.028)	0.241*** (0.027)	0.155*** (0.024)	0.163*** (0.025)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,782	9,449	9,782	9,449	9,782	9,449
School-grades	53,867	52,037	53,867	52,037	53,867	52,037
Observations	389,775	368,433	389,775	368,433	389,775	368,433
Adjusted R^2	0.430	0.432	0.393	0.406	0.336	0.341

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.9: Placebo tests pre-treatment, including year 2010 - Early dropout rate $t+1$

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated \times FakeFRP ₂₀₀₅	0.368*** (0.067)	0.358*** (0.067)	0.240*** (0.043)	0.207*** (0.043)	0.128*** (0.034)	0.151*** (0.034)
Treated \times FakeFRP ₂₀₀₆	0.330*** (0.055)	0.319*** (0.056)	0.217*** (0.035)	0.186*** (0.035)	0.113*** (0.027)	0.133*** (0.028)
Treated \times FakeFRP ₂₀₀₇	0.362*** (0.052)	0.319*** (0.052)	0.236*** (0.033)	0.190*** (0.033)	0.126*** (0.026)	0.129*** (0.026)
Treated \times FakeFRP ₂₀₀₈	0.459*** (0.053)	0.444*** (0.054)	0.303*** (0.033)	0.270*** (0.034)	0.156*** (0.027)	0.174*** (0.027)
Treated \times FakeFRP ₂₀₀₉	0.861*** (0.062)	0.836*** (0.063)	0.496*** (0.039)	0.463*** (0.040)	0.365*** (0.031)	0.373*** (0.031)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	53,376	51,401	53,376	51,401	53,376	51,401
Observations	241,505	226,760	241,505	226,760	241,505	226,760

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This finding would pose a serious threat to our identification strategy, since it raises concerns about other potential confounding factors affecting early dropout rates in the pre-treatment period. However, since we suspect an anticipatory effect in early dropout rates in 2010, as suggested by the previous common trend tests, we run once more the same placebo tests of Table 3.9 excluding year 2010 from the sample. As shown in Table 3.10, once we exclude the year of the anticipatory effect, all the estimates from the placebo tests are not significantly different from zero. This result not only confirms the required pre-treatment common trend but also confirms the occurrence of an anticipatory effect in early dropout rates in year 2010. Finally, in Table 3.11 we present the outcomes of the two possible placebo tests for the after-reform period (2011-2013), which also give non-significant estimates.

Table 3.10: Placebo tests pre-treatment, excluding year 2010 - Early dropout rate t_{+1}

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated \times FakeFRP ₂₀₀₅	0.060 (0.100)	0.065 (0.101)	0.002 (0.064)	0.005 (0.064)	0.058 (0.052)	0.060 (0.053)
Treated \times FakeFRP ₂₀₀₆	-0.019 (0.106)	0.021 (0.109)	-0.034 (0.068)	-0.022 (0.069)	0.015 (0.056)	0.042 (0.057)
Treated \times FakeFRP ₂₀₀₇	-0.074 (0.107)	-0.100 (0.108)	-0.045 (0.068)	-0.048 (0.069)	-0.029 (0.056)	-0.052 (0.057)
Treated \times FakeFRP ₂₀₀₈	0.014 (0.095)	-0.001 (0.098)	0.057 (0.061)	0.048 (0.062)	-0.043 (0.049)	-0.049 (0.051)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	52,325	50,952	52,325	50,952	52,325	50,952
Observations	196,020	183,726	196,020	183,726	196,020	183,726

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.11: Placebo tests FRP period - Early dropout rate t_{+1}

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated \times FakeFRP ₂₀₁₁	-0.21 (0.144)	-0.174 (0.147)	-0.107 (0.091)	-0.078 (0.092)	-0.103 (0.074)	-0.096 (0.076)
Treated \times FakeFRP ₂₀₁₂	0.21 (0.144)	0.174 (0.147)	0.107 (0.091)	0.078 (0.092)	0.103 (0.074)	0.096 (0.076)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	53,396	51,273	53,396	51,273	53,396	51,273
Observations	144,827	137,087	144,827	137,087	144,827	137,087

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Given the results of our common trend and placebo analyses presented before, we consider that the sizable, robust and significant anticipatory effect in early dropout rates in 2010 needs to be excluded from our main difference-in-differences model (3.2) to be able to estimate correctly the effect of the FRP reform on early dropout rates. We proceed accordingly and show the main results in the next Section.

3.6.4 Early dropout rates: The effect of the FRP reform and increased retention

Table 3.12 presents our baseline estimates of the effect of higher retention, as induced by the FRP reform, on early dropout rates at secondary school in the year following retention. As indicated in Section 4, these estimations are also weighted by the total number of students per year and include year-fixed effects, school-grade fixed effects, school-grade specific linear trends and standard errors clustered at the school-grade level.

The results in Table 3.12 provide evidence of a significant positive effect of the FRP reform on the average early dropout rate of secondary school students. Nevertheless, the effect of increased retention on early dropout seems to be rather small: on average, the dropout rate before the end of the school year in treated schools was 0.6 of a percentage point larger after the reform in comparison with control schools. This increase in dropouts was driven relatively equally by male and female students.

Table 3.12: Effect of the FRP reform on dropout rates t_{+1}

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated \times FRP $_{t-1}$	0.621*** (0.051)	0.561*** (0.052)	0.335*** (0.031)	0.296*** (0.032)	0.287*** (0.026)	0.265*** (0.026)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	53,720	51,759	53,720	51,759	53,720	51,759
Observations	344,797	325,689	344,797	325,689	344,797	325,689
Adjusted R^2	0.376	0.378	0.346	0.360	0.285	0.292

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Rescaling the estimates of Table 3.12 by the effect of the FRP reform on retention rates (see Table 3.A2 in the Appendix) suggests that a 10 percentage points rise in retention rates will increase early dropout rates by 0.8 of a percentage point in the year following retention. The effect seems to be similar between males and female students; a 10 percentage point increase in male (female) students' retention increases male (female) dropouts before the completion of the school year by 0.75 (0.83) of a percentage point.

To the best of our knowledge, there was not any other educational reform or regulation change around the same FRP period that could have affected dropout rates. Therefore, we discard the possibility that difference in early dropout rates between treated and control schools after the FRP reform are caused by any other reason than the increase in students' retention. This finding is consistent with the previous empirical literature, as summarised in Section 3.2, that suggest that grade retention leads to higher school dropout rates. The size of the effect, however, seems to be initially small in economic terms, which could suggest that the undesirable effects of retention on dropouts could be in some cases offset by its positive effects, for instance, the effect of better school performance.

3.6.4.1 Heterogeneous effects by treatment intensity

To provide further insights into the effects of retention on end-of-year dropouts at secondary school, we also analyse whether the estimates in Table 3.12 differ across schools that experience different intensity in the treatment. Following the same steps as in Section 3.6.2, we now re-estimate equation (3.2) using the heterogeneous treatment variable instead of the binary indicator and show the results in Table 3.13.

The results in Table 3.13 confirm our main finding of a significant positive effect of the FRP reform on early dropouts at secondary school. Moreover, Table 3.13 suggests this effect to be linear or proportional to the extent of retention growth: on average, the dropout rates before the end of the school year in high-treated, mid-treated and low-treated schools were respectively 1, 0.4 and 0.2 percentage points higher after the reform in comparison with control schools. These positive effects of retention on dropping out of school were slightly larger among male than female students, particularly in the case of low-treated schools.

Table 3.13: Effect of the FRP reform on dropout rates t_{+1} - heterogeneous treatment

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
Low-treated \times FRP $_{t-1}$	0.186*** (0.068)	0.207*** (0.069)	0.152*** (0.041)	0.162*** (0.042)	0.034 (0.036)	0.045 (0.036)
Mid-treated \times FRP $_{t-1}$	0.381*** (0.071)	0.368*** (0.073)	0.237*** (0.043)	0.225*** (0.044)	0.145*** (0.039)	0.143*** (0.039)
High-treated \times FRP $_{t-1}$	1.004*** (0.072)	0.934*** (0.074)	0.572*** (0.044)	0.527*** (0.045)	0.432*** (0.037)	0.407*** (0.037)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	53,720	51,759	53,720	51,759	53,720	51,759
Observations	344,797	325,689	344,797	325,689	344,797	325,689
Adjusted R^2	0.376	0.378	0.346	0.360	0.285	0.292

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As previously mentioned, Table 3.A3 in the Appendix shows that the FRP reform increased retention in low-treated, mid-treated and high-treated schools by 1.6, 4.4, and 11.1 percentage points, respectively. Among male students, the same effects were 0.9, 2.3, and 6.3 percentage points whereas for female students' retention increased by 0.7, 2.0 and 4.9 percentage points respectively. Thus, rescaling the estimates obtained in Table 3.13 to the corresponding increases in retention, reinforces the conclusion from the previous section that an increase of one percentage point in retention rises early dropout rates linearly by approximately 0.09 percentage point in the year following retention. This means, the strongest effect on dropout rates was experienced among high-treated schools.

The effect of retention on dropouts, however, seems to differ between male and female students when taking into consideration the heterogeneous treatment intensity across schools. Our findings suggest that the slight increase in retention rates among all low-treated schools only affected male early dropouts and did not have any effect on female students. Similarly, whereas a one percentage point higher retention amongst mid-treated schools led to an increase of one percentage point in male dropout rates, female dropout rates only increased by 0.06 percentage points. Nevertheless, amongst high-treated schools, the scaled effect of retention on early dropouts is similar for male and female students. Although the effect of the FRP reform on male dropout rates was slightly higher (0.6 percentage point) than for females (0.4 percentage point), this effect is proportional to the higher increase in males' retention (6.3 percentage points) in comparison with the increase in female retention rates (4.9 percentage points).

3.6.4.2 Heterogeneous effects by grade

In this last section, we test for potential differences in the effect of retention on early dropout rates across grades. For instance, it could be that retention at early grades affects more strongly students' attitudes and future perspectives than retention at later stages, which could lead to heterogeneous motivations and responses in terms of dropping out of school. To test for this hypothesis, we re-estimate equation (3.2) for each grade separately, from grade 6 to grade 11, using both the binary and the heterogeneous treatment variables. The results are shown in Table 3.14 and Table 3.15, respectively.

First, when using the binary treatment indicator as shown in Table 3.14, we find that indeed the overall positive effect of the FRP reform on early dropout rates is stronger the earlier the grade, and that there is not any significant effect during the last two years of secondary education, i.e. grades 10 and 11. Rescaling these estimates to the increase in retention due to the FRP reform (See Table 3.A4 in the Appendix) confirms that the effect of our interest is strongest at grade 6; a 10 percentage points increase in students' retention during the first year of secondary school increases early dropouts at grade 6, one year after, by 1.1 percentage points. This effects gradually decreases over grades until grade 9, when a 10 percentage points increase in students' retention rises early dropouts by 0.6 of a percentage point. Moreover, we observe in Table 3.14 that the effects of the FRP reform are, in all grades, consistently higher for male students than for female students, nevertheless, proportional to the larger increases in males'

retention across all grades as well.

Second, when using the low-, mid-, and high-treatment variables, as shown in Table 3.15, we observe large heterogeneity in the effect of the FRP reform across grades of schools treated with different intensity. For instance, the slight increase in retention amongst low-treated schools only had a marginal positive effect on early dropouts in grade 6, while all other grades seemed to be unaffected. Within mid-treated schools, the positive effect of the reform on students' dropouts appears to be significant only from grade 6 to grade 8. Students in grades 9 and 10 did not seem to react to the law change.

Surprisingly, the results for grade 11 of mid-treated schools suggest a small marginal negative effect on dropouts. Finally, the positive effect on high-treated schools is confirmed to be the largest in all grades, except grades 10 and 11, where the increase in retention induced by the reform did not affect students' early dropouts. This effect across high-treated schools is also greater the earlier the grade.

More interestingly, some gender differences become evident. Table 3.15 and Table 3.A5 in the Appendix reveal that amongst low-treated schools, for example, the small positive effect of retention on early dropouts in grade 6 is driven entirely by male students, although the increase in retention was very similar for both males and females. Within mid-treated schools, the effect of retention on dropping out of school in grades 7 and 8 is almost the same between males and females, even though males' retention rates were approximately one percentage point higher among males. Conversely, at grade 6 the same effect is twice the size for male students in comparison with female students; nonetheless, this seems to be proportional to retention rates, which affected males more strongly than females in grade 6. In the case of high-treated schools, the effects vary as well. For students in grade 9, the reform affected male and female dropouts in a very similar magnitude, even though again males' retention was approximately one percentage point higher than females' retention. The effect on students' dropouts from grade 6 to grade 8, however, appears to be proportional to the larger retention rates among male students in comparison to females after the FRP reform. Finally, there is a striking small negative effect of the reform on dropouts that is significant for grade 11 males in low- and mid-treated schools only. This is surprising since the reform, according to Table 3.A5 in the Appendix, did not have any effect on retention for this group of students. This might have been a clever response from students in male-only schools that expected correctly the FRP reform to affect students' retention in other grades except for grade 11, given the historically low retention rates at this level, so their incentives to finish the last year of secondary school could have marginally increased.

Table 3.14: Effect of the FRP reform on dropout rates t_{+1} by grade

	grade 6		grade 7		grade 8		grade 9		grade 10		grade 11	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>All</i>												
Treated \times FRP $_{t-1}$	1.046*** (0.138)	0.968*** (0.140)	0.656*** (0.117)	0.603*** (0.119)	0.654*** (0.119)	0.580*** (0.121)	0.422*** (0.114)	0.381*** (0.117)	0.061 (0.261)	0.153 (0.263)	-0.169 (0.185)	-0.087 (0.179)
<i>Males</i>												
Treated \times FRP $_{t-1}$	0.654*** (0.088)	0.596*** (0.090)	0.375*** (0.072)	0.350*** (0.073)	0.336*** (0.070)	0.313*** (0.071)	0.224*** (0.068)	0.205*** (0.070)	0.047 (0.163)	0.135 (0.163)	-0.048 (0.111)	0.006 (0.108)
<i>Females</i>												
Treated \times FRP $_{t-1}$	0.391*** (0.066)	0.372*** (0.067)	0.280*** (0.058)	0.253*** (0.059)	0.318*** (0.063)	0.267*** (0.063)	0.198*** (0.062)	0.176** (0.063)	0.014 (0.136)	0.019 (0.139)	-0.121 (0.103)	-0.093 (0.101)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,731	9,353	9,664	9,292	9,531	9,174	9,327	8,976	7,839	7,576	7,628	7,388
Observations	64,330	60,058	63,097	59,325	61,377	57,964	59,202	56,085	49,440	47,087	47,351	45,170

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.15: Effect of the FRP reform on dropout rates $t+1$ by grade - heterogeneous treatment

	grade 6		grade 7		grade 8		grade 9		grade 10		grade 11	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>All</i>												
Low-treated \times FRP $_{t-1}$	0.380** (0.181)	0.327* (0.185)	0.172 (0.161)	0.204 (0.164)	0.160 (0.163)	0.158 (0.166)	0.058 (0.152)	0.085 (0.156)	0.207 (0.347)	0.239 (0.355)	-0.307 (0.247)	-0.303 (0.241)
Mid-treated \times FRP $_{t-1}$	0.813*** (0.195)	0.704*** (0.202)	0.379** (0.160)	0.402** (0.164)	0.430** (0.172)	0.402** (0.176)	0.148 (0.157)	0.139 (0.160)	-0.113 (0.375)	0.040 (0.379)	-0.502* (0.255)	-0.459* (0.235)
High-treated \times FRP $_{t-1}$	1.622*** (0.197)	1.380*** (0.202)	1.051*** (0.168)	0.996*** (0.171)	0.997*** (0.169)	0.904*** (0.174)	0.695*** (0.157)	0.660*** (0.164)	0.416 (0.363)	0.493 (0.365)	-0.219 (0.253)	-0.127 (0.247)
<i>Males</i>												
Low-treated \times FRP $_{t-1}$	0.323*** (0.115)	0.282** (0.119)	0.147 (0.099)	0.164 (0.100)	0.148 (0.094)	0.147 (0.096)	0.028 (0.093)	0.035 (0.096)	0.272 (0.220)	0.281 (0.224)	-0.301** (0.149)	-0.302** (0.146)
Mid-treated \times FRP $_{t-1}$	0.575*** (0.120)	0.465*** (0.124)	0.201** (0.096)	0.200** (0.098)	0.225** (0.096)	0.211** (0.099)	0.043 (0.095)	0.038 (0.098)	0.098 (0.237)	0.219 (0.237)	-0.352** (0.149)	-0.295** (0.143)
High-treated \times FRP $_{t-1}$	1.057*** (0.126)	0.861*** (0.129)	0.552*** (0.102)	0.522*** (0.103)	0.580*** (0.096)	0.552*** (0.099)	0.348*** (0.096)	0.322*** (0.101)	0.301 (0.226)	0.341 (0.231)	-0.137 (0.155)	-0.087 (0.151)
<i>Females</i>												
Low-treated \times FRP $_{t-1}$	0.057 (0.089)	0.044 (0.089)	0.025 (0.083)	0.040 (0.084)	0.012 (0.094)	0.011 (0.095)	0.029 (0.087)	0.050 (0.088)	-0.065 (0.186)	-0.042 (0.192)	-0.006 (0.144)	-0.001 (0.143)
Mid-treated \times FRP $_{t-1}$	0.238** (0.105)	0.239** (0.109)	0.178** (0.083)	0.182** (0.085)	0.206** (0.100)	0.192** (0.098)	0.105 (0.086)	0.102 (0.086)	-0.211 (0.199)	-0.179 (0.205)	-0.150 (0.139)	-0.165 (0.137)
High-treated \times FRP $_{t-1}$	0.564*** (0.091)	0.519*** (0.092)	0.499*** (0.085)	0.474*** (0.087)	0.417*** (0.096)	0.352*** (0.098)	0.347*** (0.084)	0.339*** (0.086)	0.115 (0.194)	0.121 (0.197)	-0.082 (0.140)	-0.039 (0.140)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,731	9,353	9,664	9,292	9,531	9,174	9,327	8,976	7,839	7,576	7,628	7,388
Observations	64,330	60,058	63,097	59,325	61,377	57,964	59,202	56,085	49,440	47,087	47,351	45,170

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.7 Conclusions

We have estimated the aggregated causal effects of grade retention on school dropout rates in secondary schools, distinguishing between (1) the end-of-the-year effect for retained students and (2) the effect in the year after retention for all pupils enrolled in each grade. We have analysed the heterogeneity of these effects depending on the timing at which retention occurs by estimating the impact of retention on dropouts at each grade of secondary education from grade 6 to grade 11.

We highlight two major findings. First, a remarkably large positive effect of grade retention on end-of-year dropout rates among retained students and a positive, however relatively small, effect of grade failure on consecutive dropout rates among all students enrolled in the year following retention. Second, we show significant heterogeneity in both effects depending on the timing of retention along the secondary education cycle: the effects of grade failure on early dropout rates are stronger when retention takes place at the earlier grades whereas the effect among retained students is strongest if retention occurs at grade 9 and grade 11, precisely the grades where successful completion entitles the students to receive the lower secondary school certificate and the high-school diploma respectively.

These results account for the endogeneity of the selection into retention by estimating a difference-in-differences model that exploits variation in retention rates induced by a policy change with respect to retention in Colombia. Until 2009, schools were each year restricted by national regulation to retain up to a maximum of 5 percent of their students. This retention restriction was ended by the Free Retention Policy (FRP) reform, which allowed schools from 2010 onwards to retain as many students as they considered appropriate.

We provide robust evidence of a significant positive effect of increased retention on the dropout rate of retained students during secondary school. The finding that approximately 40 percent of retained students do not continue with their secondary education – or at least, interrupt their studies temporarily – due to the experience of being retained is a very relevant insight that points to the undesirable effect of retention that seems to discourage retained students to continue with their human capital development.

Notably, we also show that the overall positive effect of retention on dropout rates for the full cohort of students is stronger in the earlier grades of secondary school and that there is not any significant effect during the last two years of high school. These results highlight that the strongest dropout effect takes place during the year of transition from primary to secondary education, i.e. grade 6. Conversely, the retention effect on the dropout rates of retained students is strongest if students are retained at the end of grade 9 and grade 11, suggesting that particular attention to retained students at these two specific grades is necessary since leaving school at these stages would imply the highest costs of retention due to the forgone opportunity for students to finalise either one of the two cycles of secondary education and receive the corresponding certificate.

Finally, we provide empirical evidence of the heterogeneity and (non)linearity of the main effects with regard to the treatment intensity by distinguishing be-

tween students in schools that reacted more or less to the policy reform. Although we find most of our main effects to be linear to schools' retention growth, we observe that retained students from grade 9 to 11 in mid-treated schools had greater dropout rates than similar students in high-treated schools where retention rates were even higher. This last finding suggests that students retained at later stages of secondary education in a school environment where retention is not a common outcome can cause more harm to the future perspectives of those students than if they would have been in schools where retention was more pronounced or a more likely result at the end of the school year.

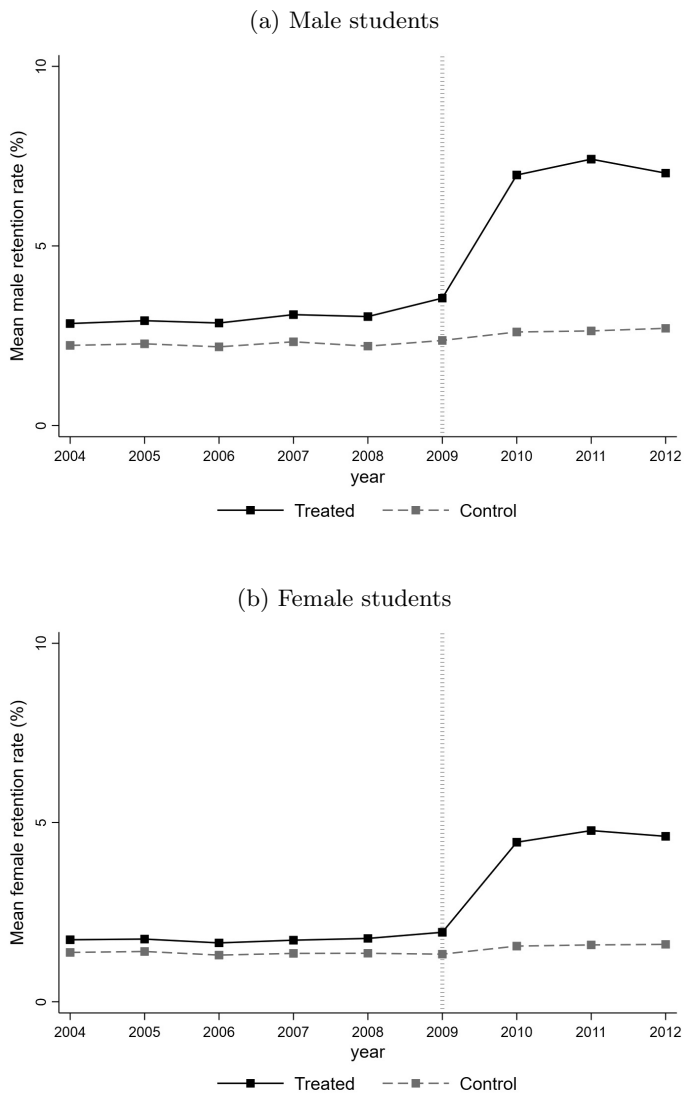
The findings of our study contribute to the evaluation of the costs and benefits of retention practices for society. It is important to take the effects of retention for retained and non-retained students into consideration, analysing both the positive and negative effects retention may have not only on school performance but also on the decisions of continuing with their secondary school studies until graduation. Because we cannot identify students who permanently leave education and those who just interrupt their studies temporarily to return later to school, our results should be interpreted cautiously as the short-term (one year) effects of retention. Further research on the long-term effects of retention on dropping out of school is then essential to unveil the full dimension of impacts that retention may have on retained students.

This study also emphasised the importance of developing a more holistic approach to assess the effects of retention. To answer the question of whether – and when – low-achieving students should be required or not to repeat a failed grade, we need to take into consideration the extent to which grade retention affects not only the retained or repeating individuals but also the extent to which grade repeaters may affect their classmates and the specifics of the context that could mitigate the undesirable effects of grade retention or intensify its positive effects on the human capital development of pupils. Our findings suggest that we must carefully assess the costs and benefits of grade retention policies for different types of students and at different moments during the school cycle. Although this study particularly focuses on the dropout costs of retention, a more holistic approach should also include the potential benefits if retention leads to better performance at exams or later in the labour market. This study has also suggested that there are differences between the processes of dropping out early and dropping out late that would be hidden by considering all dropouts together. Therefore, research applying a more holistic approach will benefit the decision-making process regarding retention policies at the country and school levels.

More accurate procedures to identify the risks and benefits of grade retention for different students in different school contexts will help schools' awareness of the extent to which they need to retain students or provide instead interventions of a different kind. These are crucial decisions knowing that retaining students is highly predictive of dropout even if not all dropouts are retained, and an important consideration when attempting to decrease dropout rates.

Appendices to Chapter 3

Figure 3.A1: Retention rates by treatment status and sex



Notes: This figure shows school average retention rates in year t for treated and control schools. Averages include grade 6 to grade 11. Panels (a) shows averages for male students and panel (b) for females students.

Figure 3.A2: Retention rates by heterogeneous treatment status

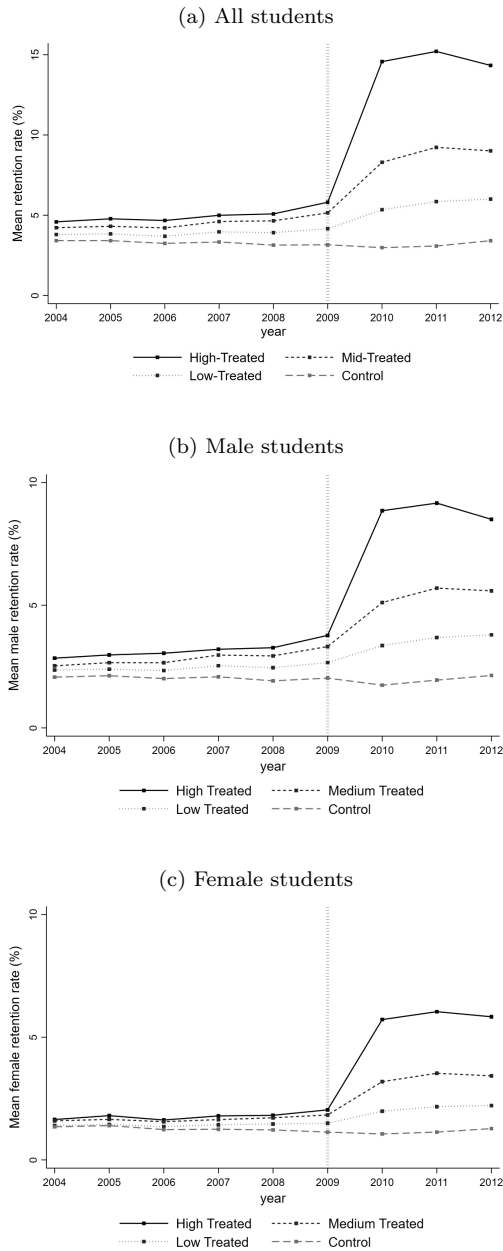


Figure 3.A3: End-of-year dropout rate ι by heterogeneous treatment status

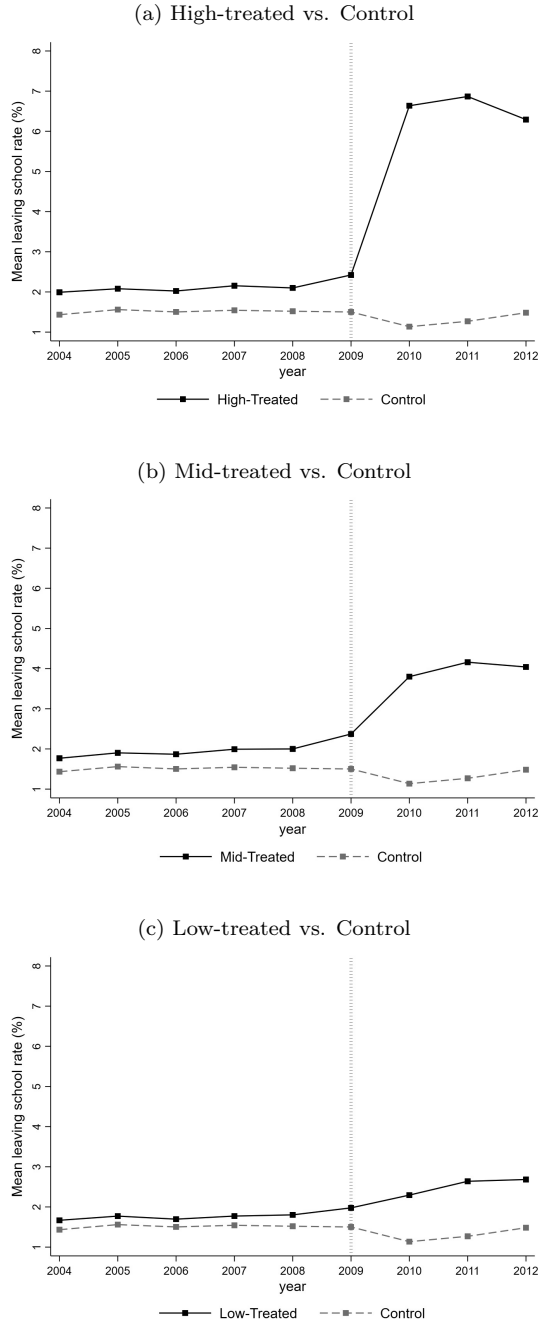


Figure 3.A4: Early dropout rates $t+1$ by treatment status and sex

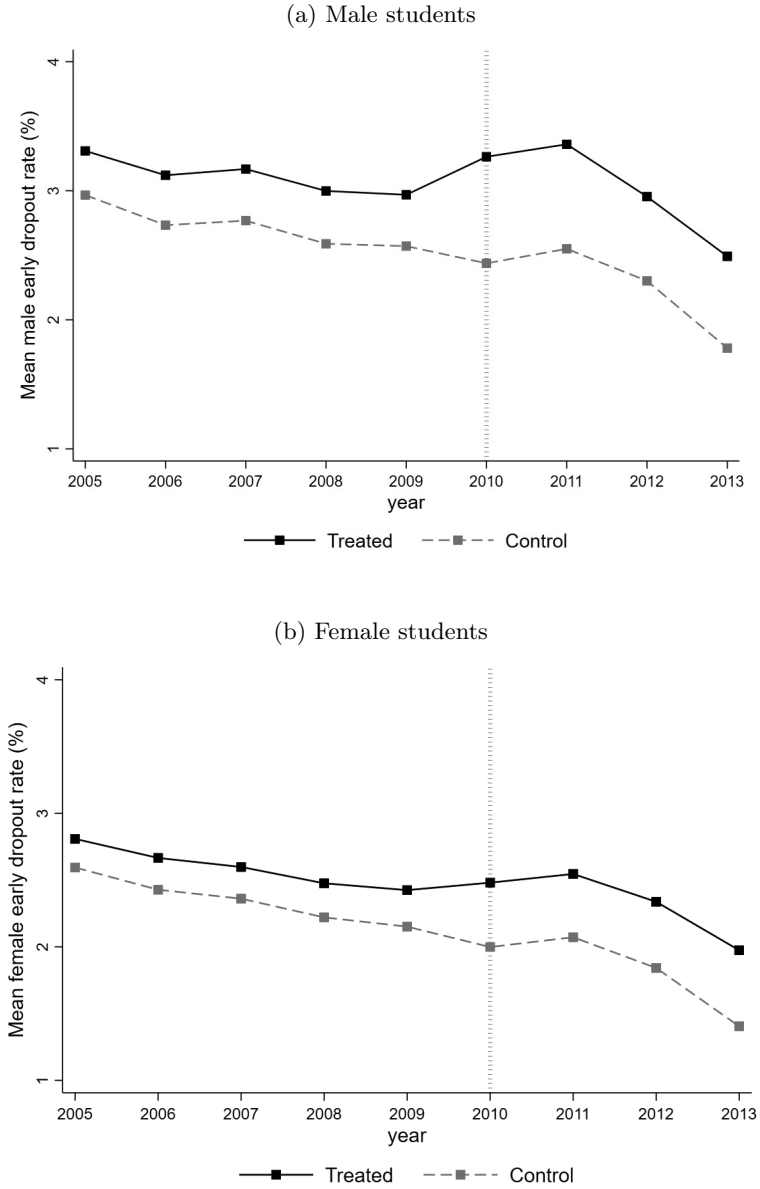


Figure 3.A5: Early dropout rate t_{+1} by heterogeneous treatment status

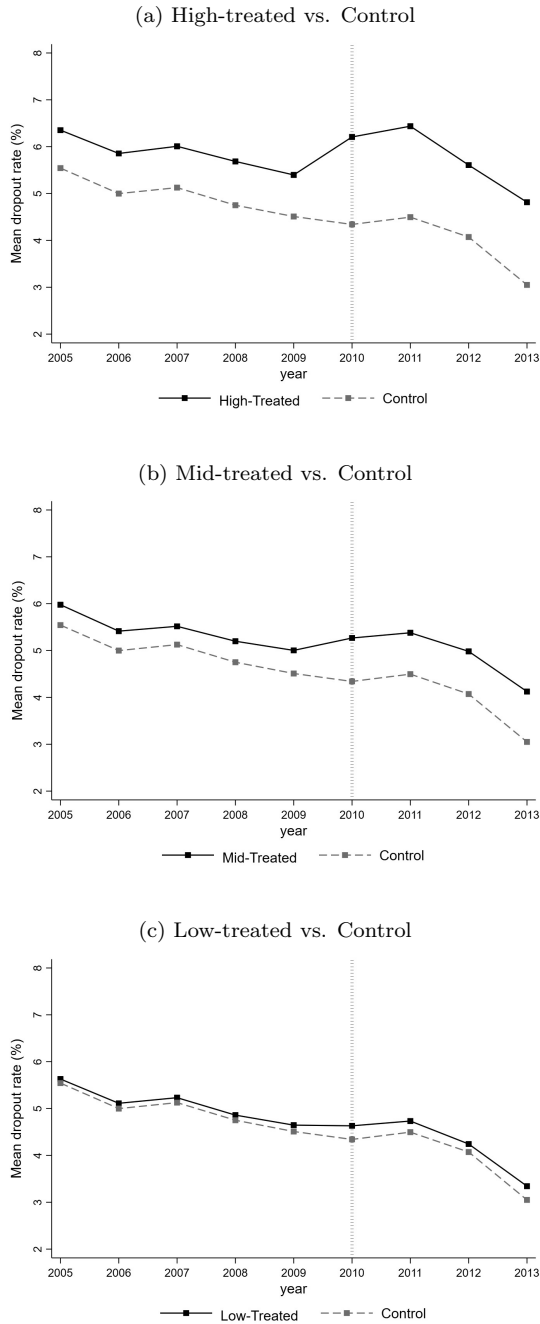


Table 3.A1: Common trend test - End-of-year dropout rate τ_t heterogenous treatment

	(1)	(2)
<i>Pre-treatment trends</i>		
Low-treated \times 2005	0.052 (0.187)	-0.148 (0.210)
Low-treated \times 2006	0.032 (0.193)	-0.115 (0.204)
Low-treated \times 2007	0.076 (0.210)	-0.135 (0.182)
Low-treated \times 2008	0.148 (0.193)	-0.019 (0.206)
Low-treated \times 2009	0.174 (0.198)	-0.029 (0.187)
Mid-treated \times 2005	-0.039 (0.187)	-0.178 (0.222)
Mid-treated \times 2006	0.196 (0.192)	-0.096 (0.217)
Mid-treated \times 2007	0.146 (0.207)	-0.128 (0.204)
Mid-treated \times 2008	0.219 (0.194)	-0.109 (0.187)
Mid-treated \times 2009	0.280 (0.205)	0.338 (0.211)
High-treated \times 2005	-0.021 (0.183)	0.036 (0.187)
High-treated \times 2006	0.038 (0.198)	0.031 (0.194)
High-treated \times 2007	-0.158 (0.212)	-0.179 (0.201)
High-treated \times 2008	0.125 (0.192)	0.067 (0.190)
High-treated \times 2009	0.501** (0.198)	0.373* (0.206)
<i>FRP trends</i>		
Low-treated \times 2010	0.985*** (0.211)	0.754*** (0.198)
Low-treated \times 2011	1.154*** (0.192)	0.897*** (0.191)
Low-treated \times 2012	1.024*** (0.220)	0.803*** (0.204)
Mid-treated \times 2010	2.350*** (0.222)	1.971*** (0.206)
Mid-treated \times 2011	2.572*** (0.209)	2.138*** (0.211)
Mid-treated \times 2012	2.403*** (0.243)	1.994*** (0.233)
High-treated \times 2010	4.976*** (0.282)	5.007*** (0.281)
High-treated \times 2011	5.006*** (0.251)	4.988*** (0.254)
High-treated \times 2012	4.404*** (0.275)	4.455*** (0.262)
Covariates	No	Yes
Schools	9,782	9,449
School-grades	53,867	52,037
Observations	389,775	368,433
Adjusted R^2	0.232	0.240

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A2: Effect of the FRP reform on retention rates and share of repeaters

	Grade retention all		Grade retention males		Grade retention females		Share of repeaters t+1	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated \times FRP	7.197*** (0.130)	7.269*** (0.135)	4.015*** (0.086)	4.050*** (0.088)	3.182*** (0.067)	3.219*** (0.069)	4.002*** (0.118)	3.980*** (0.119)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	53,720	51,759	53,720	51,759	53,720	51,759	53,720	51,759
Observations	344,797	325,689	344,797	325,689	344,797	325,689	344,797	325,689
Adjusted R^2	0.522	0.528	0.489	0.496	0.480	0.484	0.449	0.456

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A3: Effect of the FRP reform on retention rates and share of repeaters - heterogeneous treatment

	Grade retention all		Grade retention males		Grade retention females		Share of repeaters t+1	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Low-treated \times FRP	1.734*** (0.162)	1.603*** (0.163)	0.968*** (0.110)	0.866*** (0.110)	0.765*** (0.086)	0.737*** (0.088)	0.827*** (0.143)	0.903*** (0.145)
Mid-treated \times FRP	4.426*** (0.169)	4.361*** (0.170)	2.394*** (0.116)	2.332*** (0.116)	2.033*** (0.089)	2.029*** (0.091)	2.317*** (0.150)	2.318*** (0.152)
High-treated \times FRP	11.140*** (0.188)	11.139*** (0.190)	6.273*** (0.124)	6.250*** (0.124)	4.867*** (0.098)	4.889*** (0.101)	6.146*** (0.171)	6.178*** (0.171)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,777	9,409	9,777	9,409	9,777	9,409	9,777	9,409
School-grades	53,720	51,759	53,720	51,759	53,720	51,759	53,720	51,759
Observations	344,797	325,689	344,797	325,689	344,797	325,689	344,797	325,689
Adjusted R^2	0.531	0.536	0.496	0.503	0.486	0.490	0.453	0.460

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A4: Effect of the FRP reform on retention rates by grade

	grade 6		grade 7		grade 8		grade 9		grade 10		grade 11	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>All</i>												
Treated × FRP	9.112*** (0.154)	8.927*** (0.158)	8.178*** (0.145)	8.108*** (0.147)	7.667*** (0.151)	7.636*** (0.153)	6.477*** (0.146)	6.517*** (0.148)	6.691*** (0.364)	6.698*** (0.372)	2.593*** (0.253)	2.540*** (0.250)
<i>Males</i>												
Treated × FRP	5.557*** (0.106)	5.439*** (0.108)	4.582*** (0.096)	4.550*** (0.098)	4.180*** (0.098)	4.168*** (0.099)	3.555*** (0.095)	3.567*** (0.096)	3.589*** (0.236)	3.537*** (0.244)	1.439*** (0.181)	1.356*** (0.170)
<i>Females</i>												
Treated × FRP	3.555*** (0.080)	3.488*** (0.081)	3.596*** (0.078)	3.558*** (0.079)	3.487*** (0.084)	3.468*** (0.085)	2.923*** (0.082)	2.950*** (0.083)	3.102*** (0.195)	3.160*** (0.194)	1.154*** (0.120)	1.184*** (0.123)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,731	9,353	9,664	9,292	9,531	9,174	9,327	8,976	7,839	7,576	7,628	7,388
Observations	64,330	60,058	63,097	59,325	61,377	57,964	59,202	56,085	49,440	47,087	47,351	45,170

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A5: Effect of the FRP reform on retention rates by grade - heterogenous treatment

	grade 6		grade 7		grade 8		grade 9		grade 10		grade 11	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>All</i>												
Low-treated × FRP	2.64*** (0.138)	2.63*** (0.143)	2.44*** (0.127)	2.43*** (0.132)	2.29*** (0.138)	2.27*** (0.141)	2.35*** (0.134)	2.35*** (0.139)	1.08** (0.455)	0.94** (0.454)	0.26 (0.342)	0.06 (0.292)
Mid-treated × FRP	6.54*** (0.150)	6.51*** (0.154)	5.81*** (0.141)	5.79*** (0.146)	5.41*** (0.142)	5.39*** (0.143)	4.77*** (0.150)	4.81*** (0.154)	3.95*** (0.465)	3.74*** (0.466)	1.04*** (0.368)	0.88*** (0.324)
High-treated × FRP	13.82*** (0.203)	13.59*** (0.213)	12.62*** (0.187)	12.55*** (0.194)	11.87*** (0.201)	11.85*** (0.204)	10.32*** (0.196)	10.40*** (0.200)	9.92*** (0.528)	9.88*** (0.527)	4.05*** (0.392)	3.85*** (0.349)
<i>Males</i>												
Low-treated × FRP	1.72*** (0.101)	1.72*** (0.104)	1.39*** (0.094)	1.37*** (0.097)	1.29*** (0.097)	1.29*** (0.099)	1.33*** (0.094)	1.34*** (0.097)	0.48 (0.313)	0.39 (0.313)	0.12 (0.278)	-0.09 (0.207)
Mid-treated × FRP	4.12*** (0.115)	4.08*** (0.118)	3.29*** (0.107)	3.27*** (0.109)	3.02*** (0.102)	3.01*** (0.103)	2.69*** (0.105)	2.69*** (0.107)	2.01*** (0.321)	1.85*** (0.322)	0.52* (0.290)	0.31 (0.222)
High-treated × FRP	8.45*** (0.142)	8.32*** (0.148)	7.06*** (0.128)	7.03*** (0.132)	6.47*** (0.132)	6.47*** (0.134)	5.65*** (0.127)	5.70*** (0.130)	5.30*** (0.340)	5.21*** (0.340)	2.26*** (0.306)	2.04*** (0.242)
<i>Females</i>												
Low-treated × FRP	0.92*** (0.079)	0.91*** (0.081)	1.05*** (0.075)	1.06*** (0.077)	0.99*** (0.080)	0.99*** (0.081)	1.03*** (0.087)	1.01*** (0.089)	0.59** (0.246)	0.55** (0.243)	0.15 (0.142)	0.16 (0.144)
Mid-treated × FRP	2.43*** (0.090)	2.43*** (0.092)	2.51*** (0.084)	2.52*** (0.086)	2.39*** (0.091)	2.39*** (0.092)	2.08*** (0.096)	2.11*** (0.098)	1.94*** (0.250)	1.90*** (0.254)	0.53*** (0.157)	0.57*** (0.162)
High-treated × FRP	5.36*** (0.110)	5.27*** (0.112)	5.56*** (0.105)	5.52*** (0.107)	5.41*** (0.117)	5.38*** (0.119)	4.68*** (0.117)	4.70*** (0.120)	4.61*** (0.284)	4.67*** (0.286)	1.79*** (0.160)	1.81*** (0.164)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Schools	9,731	9,353	9,664	9,292	9,531	9,174	9,327	8,976	7,839	7,576	7,628	7,388
Observations	64,330	60,058	63,097	59,325	61,377	57,964	59,202	56,085	49,440	47,087	47,351	45,170

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A6: Common trend tests - Early dropout rate t_{+1} heterogenous treatment

	Dropout all		Dropout males		Dropout females	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Pre-treatment trends</i>						
Low-treated × 2006	-0.081 (0.078)	-0.017 (0.078)	-0.041 (0.050)	0.018 (0.050)	-0.040 (0.041)	-0.035 (0.041)
Low-treated × 2007	-0.046 (0.076)	-0.037 (0.077)	0.005 (0.049)	0.018 (0.049)	-0.051 (0.040)	-0.055 (0.040)
Low-treated × 2008	0.007 (0.073)	-0.002 (0.075)	0.028 (0.047)	0.028 (0.048)	-0.021 (0.038)	-0.030 (0.039)
Low-treated × 2009	-0.068 (0.073)	-0.060 (0.075)	-0.032 (0.046)	-0.038 (0.047)	-0.036 (0.037)	-0.022 (0.038)
Low-treated × 2010	0.094 (0.069)	0.077 (0.070)	0.144*** (0.044)	0.132*** (0.044)	-0.050 (0.035)	-0.055 (0.036)
Mid-treated × 2006	0.069 (0.075)	0.095 (0.078)	0.044 (0.048)	0.060 (0.049)	0.025 (0.040)	0.035 (0.041)
Mid-treated × 2007	0.071 (0.076)	0.095 (0.078)	0.058 (0.048)	0.069 (0.049)	0.014 (0.040)	0.026 (0.041)
Mid-treated × 2008	0.110 (0.076)	0.037 (0.076)	0.063 (0.048)	0.010 (0.047)	0.046 (0.040)	0.026 (0.041)
Mid-treated × 2009	0.009 (0.075)	-0.068 (0.078)	0.040 (0.047)	-0.014 (0.049)	-0.031 (0.040)	-0.054 (0.041)
Mid-treated × 2010	0.490*** (0.077)	0.444*** (0.079)	0.331*** (0.048)	0.295*** (0.049)	0.159*** (0.041)	0.149*** (0.042)
High-treated × 2006	0.000 (0.071)	-0.018 (0.072)	0.009 (0.046)	0.005 (0.046)	-0.009 (0.036)	-0.023 (0.036)
High-treated × 2007	0.048 (0.073)	0.023 (0.073)	0.062 (0.047)	0.051 (0.047)	-0.015 (0.036)	-0.028 (0.036)
High-treated × 2008	0.032 (0.076)	-0.035 (0.077)	0.026 (0.048)	-0.019 (0.048)	0.006 (0.038)	-0.016 (0.038)
High-treated × 2009	0.054 (0.074)	0.045 (0.075)	0.045 (0.047)	0.048 (0.047)	0.009 (0.036)	-0.003 (0.037)
High-treated × 2010	0.862*** (0.079)	0.809*** (0.080)	0.621*** (0.050)	0.570*** (0.051)	0.241*** (0.039)	0.239*** (0.039)
<i>FRP trends</i>						
Low-treated × 2011	0.480*** (0.069)	0.434*** (0.071)	0.355*** (0.044)	0.335*** (0.045)	0.125*** (0.036)	0.099*** (0.037)
Low-treated × 2012	0.278*** (0.065)	0.317*** (0.067)	0.162*** (0.040)	0.188*** (0.041)	0.116*** (0.035)	0.129*** (0.036)
Low-treated × 2013	0.312*** (0.066)	0.334*** (0.068)	0.184*** (0.042)	0.201*** (0.043)	0.128*** (0.036)	0.133*** (0.038)
Mid-treated × 2011	0.601*** (0.081)	0.532*** (0.083)	0.406*** (0.049)	0.357*** (0.050)	0.194*** (0.044)	0.174*** (0.045)
Mid-treated × 2012	0.396*** (0.067)	0.425*** (0.067)	0.230*** (0.041)	0.245*** (0.041)	0.166*** (0.038)	0.179*** (0.038)
Mid-treated × 2013	0.371*** (0.066)	0.400*** (0.065)	0.214*** (0.039)	0.236*** (0.038)	0.157*** (0.036)	0.164*** (0.035)
High-treated × 2011	1.157*** (0.079)	1.094*** (0.080)	0.757*** (0.048)	0.702*** (0.048)	0.400*** (0.040)	0.392*** (0.040)
High-treated × 2012	0.451*** (0.066)	0.479*** (0.066)	0.280*** (0.041)	0.294*** (0.041)	0.171*** (0.034)	0.185*** (0.035)
High-treated × 2013	0.435*** (0.064)	0.452*** (0.064)	0.272*** (0.039)	0.280*** (0.038)	0.163*** (0.033)	0.172*** (0.033)
Covariates	No	Yes	No	Yes	No	Yes
Schools	9,782	9,449	9,782	9,449	9,782	9,449
School-grades	53,867	52,037	53,867	52,037	53,867	52,037
Observations	389,775	368,433	389,775	368,433	389,775	368,433
Adjusted R^2	0.428	0.430	0.393	0.406	0.336	0.342

Notes: Robust standard errors clustered at the school-grade level are reported in parentheses. All specifications include year-fixed effects, school-grade fixed effects, and school-grade specific linear trends. The models are weighted by the total number of students per year and school-grade. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Does informal learning at work differ between temporary
and permanent workers? Evidence from 20 OECD
countries**

Abstract

This chapter analyses the relationship between temporary employment and the intensity of on-the-job informal learning across 20 developed countries. Using microdata from the OECD's PIAAC survey, we estimate an instrumented endogenous switching regression model and find that temporary employees engage in on-the-job learning more intensively than their counterparts in permanent employment. We show that this higher intensity of informal learning does not substitute for temporary workers' lower participation in formal training. Instead, both types of learning are complementary. Heterogeneous-effect analyses suggests that early career expectations of gaining a permanent contract could explain the higher informal learning investments of employees while in a temporary job.

JEL Classification: E24, J24, J41.

This chapter is joint work with Andries de Grip and Rolf van der Velden. It has been published with the same title in: *Labour Economics* 55 (2018), p. 18–40. We gratefully acknowledge comments from Lex Borghans, Bart Cockx, Denis de Crombrugge, Thomas Dohmen, Bart Golsteyn, Colin Green, Annemarie Künn-Nelen, Olivier Marie, Raymond Montizaan, Kjell Salvanes, Ludger Woessmann, Jeffrey Wooldridge, and participants of the LEER Workshop on Education Economics (2015), the Learning and Work Seminar Maastricht (2015), the IEB Workshop on Education Economics (2016), the ROA Human Capital conference (2016) and the conferences of the European Society for Population Economics (2015), the International Association for Applied Econometrics (2015), and European Association of Labour Economics (2016, 2017). We also thank three anonymous referees and a co-editor of the *Labour Economics* journal for all their insightful comments and suggestions.

4.1 Introduction

During the last two decades, temporary employment has increased substantially in many OECD countries (OECD, 2014). Ideally, on-the-job investments in human capital (i.e. training and learning from experience) in this type of jobs should improve the integration of new entrants or unemployed individuals into the labour market. However, both in public policy and in the economic literature, there is a debate about the opportunities for human capital development associated with temporary contracts (Arulampalam and Booth, 1998; Booth, Francesconi, and Frank, 2002; OECD, 2014).

Despite the debate and policy relevance, remarkably little is known about the difference between temporary and permanent employees with respect to the learning content of their jobs. Mainly due to the lack of appropriate data, the empirical literature has thus far been entirely focused on workers' training participation. In line with human capital theory, several empirical studies have provided evidence of a negative relation between temporary contracts and t[r]aining participation in various countries (e.g. Arulampalam, Booth, and Bryan, 2004; Cutuli and Guetto, 2012). The empirical question on whether and, if so, to what extent learning informally on the job differs between temporary and permanent employees still remains.

Policy makers in most OECD countries have become increasingly aware that informal learning over the working life represents also a rich source of human capital and skills development, usually recognised through the experience wage premium in the labour market (OECD, 2010, 2013). Although years of experience (or tenure) have played an important role in the economic literature as a proxy for unobservable investments in learning while working (Mincer, 1997), there are hardly any empirical studies on the learning potential of different jobs or the extent to which different workers learn from experience (Heckman, Lochner, and Cossa, 2002; Rosen, 1972; Tesluk and Jacobs, 1998).

In this chapter, we contribute to filling this gap by providing empirical evidence on the influence of temporary contracts on the intensity of informal learning at work across 20 OECD countries.¹ For our analysis, we use unique data from the OECD Programme for International Assessment of Adult Competences (PIAAC) survey conducted in 2011 and 2012. This is the first survey that provides detailed internationally comparable measures of workers' skills, job-tasks content and informal learning at work, the latter based on a conceptual framework that considers three modes of informal learning: 1) learning by doing, 2) learning from others, and 3) learning by keeping up-to-date with new products or services.

We contribute to the literature on flexible employment and human capital development in three ways. First, we estimate the extent to which the intensity of informal learning on the job differs between workers with temporary and permanent contracts. In doing so, we raise the issue of potential endogeneity of enrolment in a temporary job due to selection based on unobservable characteristics. We not only include several important and previously neglected control

¹ Table 4.1 shows the countries included in our analyses.

variables in the temporary contracts literature (such as workers' skills, learning attitude and task-job content) but also implement an endogenous switching regression model (SRM) to correct for the expected negative bias in a naïve ordinary least squares (OLS) estimation. To this end, we exploit the differential exposure of workers of different ages to employment protection legislation (EPL) and potential unemployment, using age-group-by-country information six years prior to the data generation.² We obtain consistent estimates by applying the full-information maximum likelihood (FIML) method and show our estimations to be robust. Our main results show that workers in temporary jobs invest more intensively in informal learning than their counterparts in permanent contracts do, although the former are, in line with the empirical human capital literature, less likely, on average, to participate in formal training activities.

Second, we explore the interaction between training and informal learning to analyse whether there is substitution or complementarity between these two types of learning for both temporary and permanent employees. We find evidence of a complementarity relation, regardless of the type of contract, which suggests that the higher informal learning investments of temporary workers do not substitute for the lack of formal training.

Third, we provide marginal treatment effects (MTEs) estimates to analyse the heterogeneity in workers' informal learning along the distribution of their individual unobserved characteristics. This analysis shows whether informal learning outcomes for workers on the margin of a temporary job placement change with the probability of selection into a temporary contract by marginally increasing the corresponding interaction between the unemployment rate and EPL measures we use as selection instruments in our identification strategy. Allowing for this type of heterogeneity reinforces our main finding: There is a consistent difference in on-the-job informal learning between temporary and permanent employees in favour of those who have a temporary contract. Interestingly, this difference is expected to be larger among workers with lower propensities of selection into temporary jobs, that is, those who are likely to have better unobservable characteristics (e.g. ability and motivation).

We also provide additional insights on the possible mechanisms that could explain our main result. To that aim, we perform several heterogeneous-effect analyses taking into consideration different observable individual and job-task content characteristics. In evaluating this heterogeneity, the most important concern is the possibility that firms select workers into temporary contracts due to the different task content of jobs and/or workers' different levels of skills beyond their educational level. This could lead to unobservable differences in learning opportunities by contract type. This kind of problem has been largely overlooked in the literature on flexible employment. We find that employees with higher levels of numeracy/literacy skills are indeed less likely to have temporary contracts. We also find that workers with jobs that involve tasks of greater skills

² We use male unemployment rates by country as a selection instrument, which vary within countries across five-year age groups. We collected these unemployment data from six years preceding the respondent's interview date to ensure that, first, the current (in sample) and past (in instrument) five-year age groups refer to the same age group but not to the same cohort of individuals and, second, that we use consistent unemployment information prior to the occurrence of the global financial crisis for all employees in our sample.

demand and flexibility have a lower probability of being selected into temporary contracts and are simultaneously more intensively engaged in informal learning. Furthermore, we find some heterogeneity with respect to workers' age and tenure.

Although the PIAAC data do not allow us to identify the particular mechanism driving our main finding, we discuss some heterogeneous-effect analyses in light of the theoretical idea that expectations of transition to permanent employment could be responsible for the stronger incentives to invest in informal learning while in a temporary job. Additional results from a different cross-country dataset at the European level provide descriptive support for this hypothesis.

The remainder of this chapter is organised as follows. Section 4.2 discusses the literature related to our research question. Section 4.3 presents our model and empirical strategy and discusses the plausibility of the identifying assumptions. Section 4.4 describes our data. Section 4.5 presents our main empirical results, and robustness and heterogeneity analyses. Section 4.6 concludes the chapter and discusses its main findings and implications.

4.2 Related Literature

Studying how temporary contracts influence decisions of investment in informal learning at work in comparison to permanent contracts relates to two strands of economics research. First, the human capital literature on training investments and, second, the stepping-stone or dead-end job effects of temporary contracts.

From a theoretical perspective, firms using temporary contracts to adjust the size of their labour force have fewer incentives to invest in the human capital of these employees because of the shorter expected amortisation period. Workers in this situation are also expected to invest less in the accumulation of firm-specific skills. If that is the case, then the pursuit of flexible production by firms could impose negative externalities not only on the skills development of their current flexible workforce but also on their long-term productive capacity due to suboptimal aggregate training investments (Arulampalam and Booth, 1998). Previous empirical studies have confirmed this negative relation between temporary work and training participation (Arulampalam et al., 2004; Atkinson, 1998; Cutuli and Guetto, 2012; O'Connell and Byrne, 2010; Steijn, Need, and Gesthuizen, 2006). Other studies have shown, however, that employers may invest in the general training of temporary employees due to the existence of labour market imperfections (Acemoglu and Pischke, 1999a) or to screen workers according to ability prior to offering a permanent contract (Autor, 2001).

The latter reference leads us to the related literature on the stepping-stone effects of temporary employment. Most research in this field has claimed that on-the-job learning (both training and informal learning) and skills development are probably the main channel through which temporary contracts can offer a path into permanent employment. The odds of transition to a permanent position are believed to increase with the improvement of human capital and the gain of productive experience while in a temporary job (Autor, 2001; Booth et al.,

2002; Cockx and Picchio, 2012; De Graaf-Zijl, van den Berg, and Heyma, 2011; Gagliarducci, 2005; Jahn and Pozzoli, 2013; Jahn and Rosholm, 2014). This strand of literature suggests that temporary employment could provide workers with further incentives for investments in on-the-job learning to improve their skills (or offset the deterioration of their human capital while unemployed) and thereby increase their chances of finding a more stable job. Nonetheless, these studies have also noted that, if temporary jobs are recurrent, the stepping-stone prospects decrease and human capital investments are expected to decline. Temporary employment could then become dead-end jobs.

However, due to the lack of appropriate data, little has been done to validate that learning on the job could be an important mechanism to explain the transition probabilities from temporary to permanent employment. Although we do not aim to test for this mobility, we contribute to this literature by analysing the difference in the intensity of informal learning on the job between workers with temporary and permanent contracts, which has not yet been empirically analysed.

In this respect, our study relates to the literature on human capital accumulation, which has explained that, in addition to training participation, workers' human capital development over the life course is also affected by informal learning on the job. Mincer (1974) claims that on-the-job informal learning probably constitutes the major human capital investment in the workplace. In the human capital literature, informal learning has mainly been seen as learning-by-doing or learning through experience. Following the seminal work of Arrow (1962) and Mincer (1974), years of experience (or tenure) have been considered as a proxy for the unobservable investments in informal learning. However, simply accumulating years of experience assumes that, for any worker, each hour of work is equally effective in improving skills (Heckman et al., 2002). Quiñones, Ford, and Teachout (1995) and Tesluk and Jacobs (1998) show that this does not hold, while Maurer and Weiss (2010) show that not everyone is good at learning from experience. Furthermore, firms offer different levels of learning opportunities and jobs differ widely in their learning potential (Heckman et al., 2002; Rosen, 1972).³ More recent empirical studies have emphasised that workers are continuously learning by doing and learning from other workers and that such knowledge spill-over has a positive effect on their productivity (De Grip, Sauerermann, and Sieben, 2016; Destré, Lévy-Garboua, and Sollogoub, 2008). In such settings, workers' human capital increases with tenure and converges towards the job's proficiency level and the firm's job-specific learning potential.

The literature on informal learning has some important implications for our study. First, it suggests that both the direct and indirect costs of investments in training are expected to be higher than those for informal learning since human capital accumulation and labour supply compete for employees' working time when investing in training, while they do not compete in case of learning-by-doing or learning from others (Heckman et al., 2002). Therefore, the analysis on

³ The heterogeneity of informal learning experiences in the workplace has been mostly studied in the management and organisations literature. This heterogeneity depends on the extent to which the firm or job offers employees opportunities to undertake challenging tasks, interact with others, organise their own work, and so forth (De Grip, 2009; Koopmans, Doornbos, and Eekelen, 2006; Marsick, Watkins, Callahan, and Volpe, 2015).

the relation between training and temporary contracts could not be extended and generalised to informal learning. Second, years of experience and tenure are imperfect variables to provide information on informal learning on the job. Thus, the question whether the intensity of learning on the job differ – and, if so, to what extent – between employees with temporary and permanent contracts still remains.

4.3 Model and Empirical Strategy

Our primary regression equation of interest is

$$IL_i = X_i\beta + \delta T_i + \mu_i \quad (4.1)$$

where IL is a continuous variable that measures the on-the-job informal learning intensity of worker i , X is a vector of covariates composed by worker and firm characteristics along with a set of country dummies, and T is a binary indicator of the type of contract ($T = 1$ for employees on temporary contracts and $T = 0$ for employees on permanent contracts). For this model, the difference in informal learning between workers with temporary and permanent contracts is measured by the estimate of δ .

However, the binary indicator of the type of contract T_i cannot be treated as exogenous because it is potentially based on individual self-selection or selection by employers. Unobservable worker characteristics such as ability and motivation (Autor, 2001; Givord and Wilner, 2015; Loh, 1994; Mincer, 1997) as well as time preferences and risk aversion (Belzil and Leonardi, 2007; Berton and Garibaldi, 2012; Mincer, 1997; Weiss, 1986) could affect both selection into a temporary job and investments in informal learning decision, resulting in biased estimates when using OLS. For instance, if the typical individual selected into temporary contracts has relatively lower ability or stronger time preferences for the present, then the OLS estimate of δ will underestimate the temporary contract effect. We might expect the bias to also be negative if employers tend to select temporary workers particularly for jobs of, for example, lower skill demand or lesser task complexity. If we feel these hypotheses are correct, then we would argue that δ underestimates the influence of temporary contracts on on-the-job informal learning.

We account for the endogeneity of temporary job selection by including several important and previously neglected control variables in the temporary contracts literature and estimating an endogenous switching regression model. Following Heckman (1978), Heckman and Vytlacil (1999), and Heckman, Tobias, and Vytlacil (2001), the more general model is the following. The potential informal learning outcomes (IL_0, IL_1) of the type of contract $T=(0\ 1)$ are assumed to depend linearly upon observable variables X and unobservables μ_i , as in equation (4.1). The temporary contract indicator is modelled as a nonlinear function of observables Z_i and unobservables v_i and is linked to the observed outcome IL_i through the latent variable T^* :

$$T_i^* = Z_i\gamma - v_i \quad (4.2)$$

$$T_i = \begin{cases} 1 & \text{if } T_i^* > 0 \\ 0 & \text{if } T_i^* \leq 0 \end{cases}$$

$$Prob(T_i = 1|Z_i) = \Phi(Z_i\gamma)$$

$$Prob(T_i = 0|Z_i) = 1 - \Phi(Z_i\gamma)$$

Consistent with our previous conjectures, the conditional independence assumption does not hold in these kinds of models. Instead, μ_i and v_i are allowed to be correlated by a coefficient ρ and assumed to be jointly normally distributed $(\mu, v_i) \sim N(0, \Sigma)$ (Greene, 2012; Maddala, 1986; Wooldridge, 2010). Under these assumptions, the bias caused by the correlation of the type of contract T with omitted variables is addressed by the nonzero expectation of the error term μ_i in equation (1), as follows:

$$E(IL_i|T_i = 1, X_i, Z_i) = X_i\beta + \delta + \rho\sigma_\mu \left[\frac{\phi(-Z\gamma)}{\Phi(-Z\gamma)} \right] \quad (4.3)$$

$$E(IL_i|T_i = 0, X_i, Z_i) = X_i\beta + \rho\sigma_\mu \left[\frac{-\phi(-Z\gamma)}{1 - \Phi(-Z\gamma)} \right]$$

Then, the expected difference in informal learning between temporary and permanent employees is

$$E(IL_i|T_i = 1, X_i, Z_i) - E(IL_i|T_i = 0, X_i, Z_i) = \delta + \rho\sigma_\mu \left[\frac{\phi_i}{\Phi_i(1 - \Phi_i)} \right] \quad (4.4)$$

where $\phi_i = \phi(-Z\gamma)$ and $\Phi_i = \Phi(-Z\gamma)$ are the standardised normal density and distribution functions respectively.

The model is identified through exclusion restrictions: first, including at least one instrumental variable in Z that is excluded from the IL equation (4.1) and, second, the nonlinearity of the selection equation (4.2) that provides the temporary contract propensity score $p_i(Z_i)$ and thus the correlation between μ_i and v_i . For the former, we exploit variation between and within countries in workers' differential exposure to potential unemployment by using the corresponding statistics on unemployment rate by country and age groups as measured six years prior to the survey. Identification thus first requires that the unemployment rate at the age-group-by-country level induces variation in the probability of having a temporary contract, conditional on all the other covariates. This relation indeed exists in our application (see Section 4.4.2.4 and Panel B of Table 4.4). Second, the selection instrument (the unemployment rate at $t-6$) should be independent of the unobserved component in the IL equation (4.1) and should

not directly affect the intensity of informal learning at work, conditional on the observed characteristics X . We establish the admissibility of our selection instrument in Sections 4.4 and 4.5.

To yield consistent and efficient estimates of the ATE, we apply the FIML method that simultaneously fit equations (4.1) and (4.2) of our model by minimising the determinant of the covariance matrix associated with the residuals of the reduced form of the equation system.

Furthermore, by allowing $\beta_0 \neq \beta_1$ and/or $\sigma_0^2 \neq \sigma_1^2$ and $\rho_0 \neq \rho_1$ where σ^2 represents the separate variance of μ_i in Σ and ρ , or the separate correlation parameters for permanent and temporary workers, respectively, we can obtain the interacted endogenous SRM in which the impact of observed and/or unobserved characteristics is allowed to vary between workers with permanent and temporary contracts (Heckman and Vytlacil, 2005; Maddala, 1986; Wooldridge, 2010). Then, model (4.1) becomes

$$IL_i = X_i\beta_0 + \delta T_i + T_i(X_i - \bar{X})\beta_1 + \mu_{i0} + T_i(\mu_{i1} - \mu_{i0}) \quad (4.5)$$

Since the type of contract may create interaction effects with X_i and μ_i , heterogeneous effects of temporary employment on investments in informal learning on the job may result from both observed (differences between $X_i\beta_0$ and $X_i\beta_1$) and unobserved personal characteristics (differences between μ_{i1} and μ_{i0}). This way of expressing the model emphasises our primary interest in δ , although $\delta + (X_i - \bar{X})\beta_1$ is of interest for studying how the ATE changes as a function of observables, that is, to estimate nonconstant treatment effects and ATT (Wooldridge, 2010). The model also allows one to estimate treatment effect heterogeneity as a function of unobservables U_D or MTEs (Heckman and Vytlacil, 2001, 2005, 1999; Quandt, 1972). This literature shows that the selection probability into temporary contracts given by equation (4.2), which is the propensity score $p_i(Z_i)$, is a valid instrument, given selection on unobservables, and it can be used to identify MTEs.⁴ These are, in our case, the average treatment effects for workers on the margin of a temporary job placement, a margin that varies with the unemployment and EPL measures we use as instruments.

The endogenous SRM offers some advantages over the standard instrumental variable (IV) approach.⁵ First, the SRM adds more structure to account for the

⁴ The MTE is the marginal effect of a temporary contract ($T=1$) on informal learning on the job, conditional on X and the unobservables v from the selection equation (4.2). In the MTE literature, it is customary to trace out the treatment effect against the percentiles of the distribution of v , in line with the following transformation of the selection rule in equation (4.2): $T_i^* > 0$ if $\Phi(Z_i\gamma) > \Phi(v_i)$. Since $\Phi(Z_i\gamma) = p_i(Z_i)$ represents the selection probability into a temporary contract, $\Phi(v_i) = U_D$ denotes the percentiles of the distribution of the unobservable propensity of being selected into a permanent job. The condition $T_i^* > 0$ can then be rewritten as $U_D(Z_i) > v_i$. The MTE can then be estimated as the partial derivative of the conditional expectation of IL with respect to the propensity score $p_i(Z_i)$, as follows: $MTE(X_i = x, U_{D_i} = p) = \partial E\{IL_i | X_i = x, p_i(Z_i) = p\} / \partial p$. It is thus the treatment effect for individuals with observed characteristics $X_i = x$ who are at the U_D^{th} percentile of the v distribution, implying these individuals are indifferent between a permanent and a temporary contract when the propensity score $p_i(Z_i)$ equals U_D .

⁵ In the standard case of linear endogenous variables, both the SRM and 2SLS methods lead to the same estimator. When the treatment has a nonlinear effect on the outcome or the effect

binary nature of our endogenous regressor. If the nonlinear model approximates the conditional expected function of the temporary contract selection better than a linear model, the outcome estimates of informal learning investments will be much more precise than the IV estimator (Angrist and Pischke, 2008; Wooldridge, 2010). Second, a continuous selection instrument is suitable for a binary endogenous regressor, which also facilitates the estimation of MTEs (Heckman and Vytlacil, 1999; Imbens and Wooldridge, 2009). Third, it easily adapts to handle more complicated models, such as heterogeneous-effect models. The IV approach can be quite inefficient relative to the more parsimonious SRM approach in this case. Thus, by allowing the estimation of heterogeneous effects, the SRM can be applied to estimate ATE as well as ATT and MTEs (Angrist and Pischke, 2008; Heckman and Navarro-Lozano, 2004; Wooldridge, 2010). Finally, in contrast to the SRM approach, 2SLS merely provides local ATEs instead of unconditional ATEs that are more policy relevant in the context of our research question.

However, this approach, while likely more efficient than an IV approach, is less robust. The consistency of the FIML estimator of the SRM hinges on the bivariate normality assumption of μ_i and v_i ; thus, it is asymptotically efficient if the temporary contract selection equation (4.2) is correctly specified. The better the prediction of selection into a temporary job, the more precise the estimates of the informal learning outcome will be (Heckman, 1978; Heckman and Vytlacil, 2005; Wooldridge, 2010). Since the benefit of increased precision might be at the cost of a greater chance of misspecification error, we perform various robustness checks of our estimations.

4.4 Data and Descriptive Statistics

4.4.1 Data and Sample

We use data from the OECD PIAAC survey, conducted in 2011 and 2012, based on a representative sample of the population of the participant countries.⁶ This is a unique dataset that provides, for the first time, internationally comparable measures of informal learning at work, based on a conceptual framework that considers three learning pathways, namely, learning by doing, learning from others, and learning by keeping up-to-date with new products or services. Moreover, this dataset provides information on several worker, job, tasks and employer characteristics.

We restrict our sample to full-time male workers⁷ – excluding self-employed and armed forces employees – aged 17 to 65, not participating in any formal

is heterogeneous, however, the methods produce different estimates. Imbens and Wooldridge (2009) conjecture that the SRM estimator is likely much more precise because it solves the nonlinear endogeneity by adding the scalar \hat{v} to the regression while keeping the treatment variable in the second stage not only the linear projection as in 2SLS.

⁶ See OECD (2014b) for further details about data validation.

⁷ We focus on males due to the higher probability of working career interruptions among women. Temporary jobs could differ in significance between men and women, since women might prefer career flexibility throughout a significant portion of their working lives (Booth et al., 2002).

education programme, and who have an employment contract that is not an apprenticeship or other training scheme. The sample consists of 25,853 observations⁸ for 20 OECD countries,⁹ with 88.2% permanent contracts and 11.8% temporary contracts. The distribution of permanent and temporary contracts in the sample is similar and positively correlated (Pearson’s correlation = 0.79) to that distribution in the OECD statistics published for 2012 (see Table 4.1).¹⁰

Table 4.1: Sample description

Country	Sample obs.	%	Permanent	%	% OECD	Temporary	%	% OECD
1 Austria	1,269	4.9	1,191	93.9	90.7	78	6.1	9.3
2 Belgium	1,216	4.7	1,162	95.6	92.9	54	4.4	7.1
3 Canada	1,193	4.6	1,073	89.9	87.0	120	10.1	13.0
4 Czech Republic	1,193	4.6	1,029	86.3	92.6	164	13.7	7.4
5 Denmark	1,747	6.8	1,637	93.7	92.2	110	6.3	7.8
6 Estonia	1,598	6.2	1,453	90.9	95.3	145	9.1	4.7
7 Finland	1,265	4.9	1,162	91.9	87.2	103	8.1	12.8
8 France	1,682	6.5	1,536	91.3	85.6	146	8.7	14.4
9 Germany	1,361	5.3	1,225	90.0	86.1	136	10.0	13.9
10 Ireland	939	3.6	809	86.2	90.1	130	13.8	9.9
11 Italy	930	3.6	839	90.2	87.1	91	9.8	12.9
12 Japan	1,507	5.8	1,342	89.1	91.4	165	10.9	8.6
13 Korea	1,164	4.5	907	77.9	78.9	257	22.1	21.1
14 Netherlands	1,172	4.5	1,035	88.3	81.4	137	11.7	18.6
15 Norway	1,226	4.7	1,166	95.1	93.3	60	4.9	6.7
16 Poland	1,536	5.9	952	62.0	72.6	584	38.0	27.4
17 Slovak Republic	1,193	4.6	1,021	85.6	93.6	172	14.4	6.4
18 Spain	1,096	4.2	923	84.2	78.0	173	15.8	22.0
19 Sweden	1,178	4.6	1,101	93.5	85.7	77	6.5	14.3
20 United Kingdom	1,388	5.4	1,232	88.8	94.1	156	11.2	5.9
Total	25,853	100	22,795	88.2	87.8	3,058	11.8	12.2

⁸ In Canada, the sample consisted of 5,044 cases, from which we took a random sample of 1,193 cases to reduce bias due to oversampling of the Canadian respondents.

⁹ Four countries were excluded from our sample: Australia, Cyprus, the Russian Federation, and the United States. Australian data were not available due to confidentiality reasons. OECD statistics for Cyprus were not available. Data from the Russian Federation were preliminary and considered by the OECD (2014b) to not be representative of the population. Finally, the particular characteristics of the US labour market led to a loss of 58% of observations due to employees who stated not having any contract at all. In that case, only 387 nonrandom observations would have remained in our sample, of which 31.3% presumably corresponded to temporary jobs, a percentage very different from the OECD statistic, which estimates only 4.2% temporary employment in the United States. Therefore, our main variable of interest would capture something different in the United States, not comparable to other countries. As shown by the ILO (2010) and the OECD (2006), due to very low EPL, the distinction between temporary and permanent employment is of much less significance in the United States.

¹⁰ According to OECD concepts, temporary employment includes dependent workers (i.e. wage and salary workers) whose job has a predetermined termination date or both the employer and the employee understand that the duration of the job is limited. Statistics are comparable across countries, since national definitions broadly conform to this generic definition.

4.4.2 Variables

4.4.2.1 Outcome variable

On-the-job informal learning intensity is an index¹¹ derived from the following questions; all measured on a five-point Likert scale:¹²

- a) How often do you learn new work-related things from co-workers or supervisors?
- b) How often does your job involve learning-by-doing from the tasks you perform?
- c) How often does your job involve keeping up-to-date with new products or services?

This variable takes the lowest value if all three questions were answered 'never' and the highest if all were answered 'every day'. To facilitate the interpretation of results, the variable was standardised. In Table 4.2, which presents summary statistics, we observe that practically every person learns something on the job (98% informal learning incidence), with no significant difference by type of contract. Nonetheless, the mean value of informal learning intensity among temporary workers (0.07) is significantly higher than that of permanent employees (0.01). In contrast, we observe that the latter participate significantly more often in formal job-related training.

4.4.2.2 Explanatory variable

The temporary contract variable is a dummy variable that takes the value one for temporary contracts and zero for permanent contracts. Temporary contracts in our sample include fixed-term positions (90.5%) and agency work (9.5%).¹³

4.4.2.3 Control variables

As suggested by the empirical human capital literature, we control for age, educational level (years of education proxied by the highest level of education obtained), educational mismatch (dummies for overeducation and undereducation),¹⁴ firm tenure, actual weekly working hours, learning attitude,¹⁵ numeracy

¹¹ This index was derived by the OECD using the generalised partial credit model estimated by weighted likelihood. Its validity was assessed based on cross-country comparability, scale reliability and scale correlations (OECD 2014b). Our findings are robust to different constructions of the index; for example, very similar results are obtained when using the standardised principal component factor of the three statements.

¹² The response rate to these questions was 98%, with the following answer options: 1) never, 2) less than once a month, 3) less than once a week but at least once a month, 4) at least once a week but not every day, and 5) every day.

¹³ The variable is derived from the survey question what kind of employment contract do you have? The answer options were 1) an indefinite contract, 2) a fixed-term contract, 3) a temporary employment agency contract, 4) an apprenticeship or other training scheme, and 5) no contract.

¹⁴ These dummies are derived from the following question: Thinking about whether this qualification is necessary for doing your job satisfactorily, which of the following statements would be truest? The answer options were 1) this level is necessary, 2) a lower level would be sufficient, and 3) a higher level would be needed.

¹⁵ This variable is a standardised index derived by the OECD (labelled learning readiness) from the following questions, all measured on a five-point Likert scale: 1) When I hear or

test score as measured by the PIACC-OECD test, firm size (five categories), occupation (nine ISCO one-digit categories), industry (21 ISIC one-digit categories), year of interview (dummy), and country fixed effects.

Table 4.2 shows that temporary employees in our sample are generally younger and have fewer years of work experience and tenure than permanent workers. The average test scores in both numeracy and literacy are also shown to be lower among temporary workers. Moreover, among individuals in temporary positions, there is a higher share of overeducated workers and a lower proportion employed in skilled occupations, large firms, and the tertiary sector of the economy. It is worth mentioning that there is no descriptive difference between permanent and temporary employees regarding years of education, undereducation, learning attitude, and working hours.

4.4.2.4 Selection instrumental variable

To exploit workers' differential exposure to potential unemployment, we use as a selection instrument in our model the annual male unemployment rate six years preceding the respondent's interview date,¹⁶ which varies both between countries and within countries across five-year age groups. We collected these data from the OECD Statistics website.

The unemployment rate is a relevant selection instrument for the individual probability of having a temporary contract. Unemployment measures have been shown to be correlated with the subsequent incidence of temporary employment (Holmlund and Storrie, 2002; Virtanen, Kivimäki, Joensuu, Virtanen, Elovainio, and Vahtera, 2005; Wasmer, 1999).¹⁷

The average likelihood that workers will be employed in temporary jobs increases with relatively high unemployment rates. This is expected, since temporary jobs have been promoted as a mechanism to improve the labour market integration of the unemployed (Gagliarducci, 2005; Gebel, 2013) and because a

read about new ideas, I try to relate them to real-life situations to which they might apply; 2) I like learning new things; 3) when I come across something new, I try to relate it to what I already know; 4) I like to get to the bottom of difficult things; 5) I like to figure out how different ideas fit together; and 6) if I don't understand something, I look for additional information to make it clearer. This variable was constructed by the OECD based on the work of Kirby et al. (2003). Learning attitude describes workers' interests in learning and information-processing strategies and it is considered a metacognitive ability that structures the learning process and affects learning efficiency.

¹⁶ We use unemployment data from year $t-6$ to ensure that, first, the current and past five-year age groups are equivalent but not the same cohort of individuals in our sample and, second, that this unemployment information is measured before the financial crisis. Therefore, unemployment rates for 2005 and 2006 were correspondingly used for individuals interviewed in 2011 and 2012. Since we exploit unemployment variation across specific labour market segments rather than changes in unemployment over time, our results are robust to the use of unemployment data before 2005. For example, similar results are obtained when using unemployment data from 2001 to 2004.

¹⁷ Transitions from unemployment to temporary or permanent employment and from temporary to permanent contracts are likely to depend on the labour market's structure. If the pool of unemployed is large, transition rates from unemployment into temporary jobs are higher than the flow from unemployment to permanent jobs (Holmlund and Storrie, 2002; Wasmer, 1999). Similarly, if the unemployment rate is high, the probability that a temporary contract is converted into a permanent contract is lower (Givord and Wilner, 2015; Wasmer, 1999).

Table 4.2: Summary statistics

	Permanent		Temporary		Diff. p-value	Min	All	Max
	Mean	Std. Dev.	Mean	Std. Dev.				
Informal learning incidence	0.98	0.14	0.97	0.17	0.254	0	1	
Informal learning intensity (original scale)	3.40	1.09	3.51	1.18	0.001	1	5	
Informal learning intensity (standardised)	0.01	0.95	0.07	1.06	0.001	-3.28	2.05	
Training (participation)*	0.52	0.50	0.39	0.49	0.000	0	1	
Missed training due to a random event*	0.24	0.15	0.24	0.15	0.986	0	1	
<i>Age</i>	42.12	11.11	35.95	12.78	0	17	65	
Years of education	13.3	2.89	12.92	3.09	0.167	3	22	
Work experience (years)	21.37	11.68	14.58	12.54	0	0	47	
Overeducated	0.23	0.42	0.3	0.46	0.005	0	1	
Undereducated	0.07	0.26	0.06	0.23	0.111	0	1	
Learning attitude (standardised)	0	0.99	-0.02	1.09	0.221	-6.81	8.81	
Tenure (years)	11.9	10.26	4.43	7.3	0	0	45	
Weekly working hours	42.53	7.29	42.58	8.36	0.757	28	60	
Numeracy skills score	285.73	47.99	270.16	49.29	0	90.41	414.02	
Literacy skills score	281.4	43.67	271.42	46.59	0	85.82	436.59	
Interviewed in 2012	0.4	0.49	0.36	0.48	0.064	0	1	
<i>Occupation</i>								
Managers	0.1	0.3	0.05	0.21	0	0	1	
Professionals	0.18	0.38	0.15	0.35	0	0	1	
Technicians and associate professionals	0.18	0.39	0.11	0.31	0	0	1	
Clerical support workers	0.07	0.25	0.08	0.27	0.08	0	1	
Services and sales workers	0.09	0.28	0.11	0.32	0.001	0	1	
Skilled agricultural and fishery workers	0.01	0.1	0.02	0.13	0.003	0	1	
Craft and related trades workers	0.18	0.39	0.22	0.41	0	0	1	
Plant and machine operators	0.13	0.34	0.17	0.37	0	0	1	
Elementary occupations	0.05	0.22	0.11	0.31	0	0	1	
<i>Industry</i>								
Agriculture, forestry and fishing	0.02	0.13	0.03	0.16	0.056	0	1	
Manufacturing	0.3	0.46	0.3	0.46	0.358	0	1	
Construction	0.11	0.31	0.14	0.34	0	0	1	
Sales, transport, and accommodation	0.22	0.41	0.2	0.4	0.039	0	1	
Information and communication	0.05	0.21	0.02	0.16	0	0	1	
Finance	0.03	0.17	0.02	0.13	0.002	0	1	
Real estate	0.01	0.09	0.01	0.1	0.138	0	1	
Professional, technical and administration services	0.08	0.26	0.09	0.28	0.051	0	1	
Public administration, education and health	0.17	0.38	0.15	0.36	0.011	0	1	
Other services	0.03	0.17	0.04	0.2	0.049	0	1	
<i>Firm size</i>								
Firm size 1-10	0.20	0.40	0.24	0.43	0	0	1	
Firm size 11-50	0.30	0.46	0.32	0.47	0.014	0	1	
Firm size 51 -250	0.25	0.44	0.24	0.43	0.099	0	1	
Firm size 251-1000	0.14	0.35	0.12	0.33	0.005	0	1	
Firm size >1000	0.1	0.3	0.07	0.26	0	0	1	
Observations	22,795		3,058			25,853		
<i>Selection instruments</i>								
Unemployment rate $t-6$ (age-group-by-country)	6.39	4.32	10.66	7.53	0	0	0.58	
Unemployment rate $t-6$ (standardised)	-0.08	0.81	0.7	1.54	0	-1.16	10.17	
EPL regular employment (index)	2.12	0.56	2.17	0.54	0	0.92	3.05	
EPL regular employment (standardised)	0.09	0.99	0.13	0.98	0	-1.89	1.56	
EPL temporary employment (index)	1.52	1.19	1.34	1.07	0	0.05	3.93	
EPL temporary employment (standardised)	0.04	1	-0.11	0.98	0	-1.19	2.06	

*Due to lower response rate, we have fewer observations for these variables (23,232).

Table 4.3: Unemployment rate 2005/2006 and workers' personal and job-task characteristics 2011/2012

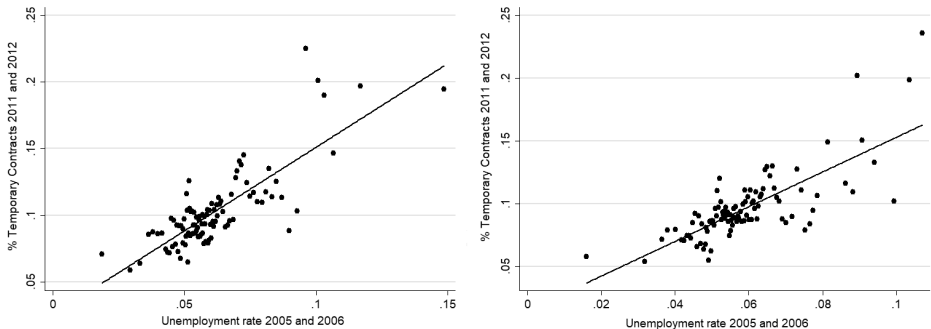
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Job satisfaction	-0.002 (0.003)	0.000 (0.003)												
Learning motivation	-0.002 (0.003)		-0.003 (0.003)											
Social trust	0.003 (0.003)			0.003 (0.003)										
Political efficacy perception	0.004 (0.003)				0.002 (0.003)									
Health status	-0.005 (0.005)					-0.004 (0.005)								
Task flexibility	0.004 (0.006)						0.000 (0.004)							
Task collaboration and teamwork	0.002 (0.004)							0.000 (0.003)						
Problem-solving tasks	-0.001 (0.004)								-0.004 (0.003)					
Planning tasks	-0.001 (0.002)									-0.001 (0.002)				
Physical tasks	0.004 (0.003)										0.002 (0.003)			
Numeracy-related tasks	0.002 (0.004)											-0.002 (0.003)		
Literacy-related tasks	-0.004 (0.004)												-0.005 (0.006)	
ICT-related tasks	-0.002 (0.005)													-0.003 (0.003)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
<i>N</i>	23,020	25,846	25,850	25,795	25,748	25,845	25,413	25,817	25,804	23,448	25,840	25,853	25,852	25,853
<i>R</i> ²	0.835	0.835	0.835	0.835	0.835	0.835	0.835	0.836	0.836	0.834	0.835	0.835	0.835	0.835

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is the dependent variable. Other controls include age and country fixed effects.

higher unemployment rate often means a risk for the active working population and job seekers that reduces the chances of finding more stable employment (European Commission, 2010a). If employment prospects are tight, workers anticipate scarce opportunities in the labour market; therefore, the probability of accepting temporary contracts is higher (Abraham, 1990; Givord and Wilner, 2015; Holmlund and Storrie, 2002). Moreover, if there is excess supply in the labour market, firms are more willing to make use of temporary contracts as a low-cost short-run buffer or as a probationary period (Gagliarducci, 2005; Givord and Wilner, 2015; Wasmer, 1999). Temporary employment then involves more employable individuals who could have had permanent contracts if the economic conditions had been better and vice versa (Givord and Wilner, 2015; Wasmer, 1999).¹⁸

Table 4.2 shows that the average value of the unemployment rate six years prior to the respondent’s interview date is three percentage points higher for the group of temporary workers than for the permanent employees in our sample. We find that the national unemployment rate by age has a significant Pearson’s correlation of 0.57 with the temporary contract dummy of our sample. Figure 4.1 shows this positive correlation, after controlling for age and country-fixed effects. To confirm that the positive correlation is not fully determined by the extremely high unemployment rates, the left-hand graph excludes unemployment rates higher than 15%, and the right-hand graph, the rates higher than 10%. In Section 4.5, we formally confirm the existence, statistical significance and robustness of this correlation in our SRM application.

Figure 4.1: Correlation graphs between unemployment rate and temporary contracts



Notes: This figure shows the correlation between the share of temporary contracts and the unemployment rate six years before, by corresponding country and age group. Each graph is a binned scatterplot of 100 bins. Linear estimation is shown, after controlling for age and country-fixed effects.

¹⁸ The greater value of hiring a worker with a temporary contract comes from the employer’s right to dismiss unproductive or mismatched workers at a lower cost. In Wasmer’s (1999) model, if the unemployment rate is relatively high, firms use more temporary contracts, and accordingly, the effect of a higher unemployment rate is to increase the share of temporary jobs.

Since the exogeneity of our selection instrument cannot be tested, it could be questioned if unemployment rates at the age-group-by-country level would affect workplace informal learning at the individual level six years later – due to a link with unobserved confounding factors. The first questioning seems to be obviously related to the type of contract an employee has: If workers are aware that the past unemployment rate was relatively high for their equivalent age group, they might face tenure uncertainty that might encourage them to invest more in on-the-job learning. However, this job tenure uncertainty mainly depends on whether the job contract is permanent or temporary.

Alternatively, the exogeneity of the unemployment rate could be questioned on the basis of an indirect link with some individual or job characteristics that are omitted from equation (4.1). This should, however, not be a concern in our setting for at least two reasons. First, we are able to take account of previously neglected control variables in the temporary contracts literature, including a large set of job task-content characteristics and some further individual characteristics.

Second, several studies on individuals' preferences have shown that adults' preferences at the individual level are highly stable and correlated over time and that any remaining instability is largely attributable to measurement error (Chuang and Schechter, 2015; De Oliveira, Croson, and Eckel, 2012; Hardardottir, 2017; Meier and Sprenger, 2015). Moreover, sporadic changes in individual preferences have been shown to be related to individuals' own experiences rather than to the use of available historical data or macroeconomic developments (Buccioli and Zarri, 2015; Malmendier and Nagel, 2011, 2016). These changes have been commonly studied in relation to extreme experiences such as a natural disaster, a violent event or an economic crisis, which have been shown to be correlated to increases in risk aversion and impatience (Callen, Isaqzadeh, Long, and Sprenger, 2014; Cameron and Shah, 2015; Cohn, Engelmann, Fehr, and Maréchal, 2015; Guiso, Sapienza, and Zingales, 2013). Individuals have also been found to be influenced more strongly by recent than distant experiences since the memory of these events vanishes over time (Malmendier and Nagel, 2011, 2016).

Drastic changes in macroeconomic variables such as unemployment could then be correlated with people's subjective preferences. However, this does not hamper our empirical strategy. First, we exploit variation in unemployment across specific labour market segments rather than unemployment shocks or changes in unemployment over time. Second, the literature suggests that aggregate measures of unemployment are not expected to be correlated to individual preferences unless an individual unemployment experience occurs (Buccioli and Zarri, 2015; Hardardottir, 2017; Malmendier and Nagel, 2011, 2016).

Third, it is important to note that the exogeneity of our instrument is further reinforced by using its lagged nature, that is, aggregate unemployment rates being measured six years preceding the survey generating our main data – which means that the current and past five-year age groups are equivalent but not the same cohort of individuals¹⁹ – and prior to the occurrence of the global financial

¹⁹ e.g. for employees in the age 35-39 interviewed in 2011 in our sample, we assign as instrument the country unemployment rate of those in the age 35-39 in 2005, when the employees

crisis.

Although we cannot formally test for the exogeneity of the unemployment rate at the age-group-by-country level, we can provide some evidence that this instrument is orthogonal to some obvious aspects such as skill levels, job satisfaction, learning motivation, or the job's tasks six years later. Table 4.3 reports the results of a battery of OLS regressions with our instrument as the dependent variable and some individual and task characteristics that could influence the opportunities and motivation for informal learning as explanatory variables. These results confirm the intuition that the unemployment rate at time $t - 6$ does not correlate significantly with any of these observable individual characteristics, whether the explanatory variables are included all together or one by one. This suggests that our selection instrument is likely to be orthogonal to other unobservable characteristics of similar nature. In principal, we could consider including the variables in Table 4.3 as control variables in our main model, but we prefer to exclude them since some of them could induce further endogeneity, whereas others could capture the impact of more relevant general variables, such as the employee's occupation or learning attitude.²⁰

An additional argument that could question the validity of our exclusion restriction relates to the technology of skill development over the life cycle (Cunha and Heckman, 2007). If acquired skills raise both the level and productivity of subsequent learning investments in life, individuals exposed to a high level of unemployment for their age group may feel pushed to invest more in schooling, thus boosting their accumulation of human capital process over their life span. This however does not invalidate our exclusion restriction since not only individuals' years of education but also their skill levels are properly accounted for in our specification. A high level of unemployment could also lead to age-targeted training policies for the unemployed. This might again lead to sub-groups of the population with a boosted accumulation of human capital. If this would not be correctly captured by the observable characteristics included in X (e.g. skill level, occupation, industry), potential differences in human capital accumulation within age groups could pose a threat to the validity of our instrument. Table 4.3 partly shows that this is not likely to be the case because, as we have explained, this argument is related to individuals' own experience of unemployment and the current and past five-year age groups do not refer to the same cohort of individuals in our setting.

Nonetheless, to provide more insight into this discussion, we perform some analyses using information on training participation among the unemployed individuals in our dataset. As shown in Table 4.A1, there does not seem to be a significant association between the contemporary (2011-2012) unemployment rates at the age-group-by-country level and the participation in (job-related) training among the unemployed individuals.

Moreover, as suggested in the literature, training programs for the unemployed in OECD countries are typically targeted on unemployment benefit recipients. Age does not usually define the main target groups and when it does,

referred were 29-33 years old.

²⁰ Including these variables, however, does not substantially change our main results. See section 4.5.4.2.

it refers to the more generic categorisation (e.g. youth, younger adults and older adults). These training programmes more often aim at a rapid reintegration to the labour market (job-search training and tackling skill deficits) instead of human capital accumulation. Studies that evaluate the effectiveness of these programs find that variation in participation is largely explained by the different institutional environments and individual characteristics and preferences, including individual unemployment duration history, educational level, family income, among others (Caliendo and Schmidl, 2016; Heckman and Navarro-Lozano, 2004; OECD, 2013). Based on this discussion, we consider our selection instrument to be credibly valid.

Finally, we estimate an alternative specification in which we use the unemployment rate interacted with the OECD EPL indexes for permanent and temporary employment as selection instrument.²¹ With this instrument we build on the expectation that the relation between unemployment and the probability of having a temporary contract may differ by country because of the strictness of EPL. Stricter rules applicable to permanent (temporary) employment may tend to increase (decrease) the incidence of temporary work and affect the extent to which temporary contracts are converted into permanent ones (Booth et al., 2002; Gagliarducci, 2005; Holmlund and Storrie, 2002; Kahn, 2010; OECD, 2004; Sala, Silva, and Toledo, 2012). In Table 4.2 we observe that, indeed, the average value of EPL applicable to permanent (temporary) employment is slightly higher (lower) for the group of temporary workers in our sample.

4.5 Empirical Results

4.5.1 On-the-Job Informal Learning Intensity

Our main results are reported in Table 4.4. The first specification gives the estimates of an ordinary OLS regression. Specifications (2) and (3) show the coefficients from standard 2SLS estimations. Specifications (4) and (5) provide the FIML results obtained from our identification strategy, that is, the endogenous SRM described in Section 4.3. Panel A of Table 4.4 shows the linear estimates of the informal learning equation and Panel B the corresponding linear/probit estimates of the temporary contract equation. In all the regressions, standard errors are clustered at the age-group-by-country level which is the exact level of variation of the selection instrument.

The results in Table 4.4 provide clear evidence of a significant difference in on-the-job informal learning between temporary and permanent employees, in favour of those who have a temporary contract. Compared with the OLS estimates, the coefficients from the estimations that account for the endogeneity of temporary contracts are adjusted upwards, in line with the negative bias we expected.

As discussed in Section 4.3, we consider the FIML estimates presented in columns (4) and (5) the preferred regressions. The main result of our study

²¹ EPL for permanent employment is a weighted indicator concerning the regulations for individual and collective dismissals. EPL for temporary employment is a weighted indicator concerning regulation on the use of fixed-term and temporary agency contracts.

is the positive and significant difference in the intensity of on-the-job informal learning between workers with temporary and permanent contracts. Specifications (4) and (5) show that, once selection into the contract type is controlled for, the estimated ATE of interest increases from 0.09 to 0.22. This result implies that workers in temporary jobs invest, on average, 0.22 of a standard deviation more in on-the-job informal learning than their counterparts in permanent employment.²² The size of this coefficient seems to be substantial if we consider that it is similar to the impact of approximately 10 years of schooling.

We consider the FIML estimates more appropriate for various reasons. First, the nonlinear prediction of selection into temporary contracts is more accurate. Whereas the linear predictions from the 2SLS first stage range from -1.68 to 0.71 (with a mean standard error of 0.09), leaving 26% of the sample below zero, the probit probabilities from the FIML estimations range from 0.01 to 0.84 (with a mean standard error of 0.02), which provides better common support. Second, the sizes of the instrument coefficients differ significantly between the 2SLS and FIML specifications. In column (4), for instance, an increase of one standard deviation in the unemployment rate, on average, increases the probability of being in a temporary contract by 1.5 percentage points. In column (2), the same effect predicted by the 2SLS is 4.8 percentage points, three times bigger. The size of the probit marginal effect is closer to that in related research, for example, Kahn (2010). Third, as we observe, the 2SLS approach inflates the estimate for temporary contracts, which might be due to the less precise prediction provided in the selection equation and because the use of a continuous instrument for the temporary contract variable makes the local ATE estimation less straightforward.²³ Moreover, as we show in Section 4.4, there is some important heterogeneity in our application, which significantly reduces the efficiency of the standard IV estimation. Fourth, we observe some implausible estimates in the 2SLS outcome equations, such as positive nonsignificant coefficients for age and tenure. Fifth, the Wald tests for specifications (4) and (5) indicate with 95% confidence that we can reject the null hypothesis of no correlation between the errors of the temporary contract and informal learning equations, so that our instrumented endogenous SRM fits well overall. The estimated correlation between the temporary contract equation errors and the outcome errors ρ is negative (-0.09), indicating that unobservables that raise informal learning tend to occur with unobservables that lower temporary contract selection. This finding is coherent with our hypothesis on the unobservables mentioned in Section 4.3; for instance, people with greater ability or lower time discount rate are less likely to be selected into temporary jobs and are at the same time more likely to invest in human capital on the job (Mincer 1997). Last but not least, in contrast to the FIML approach 2SLS does not provide ATEs but, instead, local ATEs, the former being more policy relevant in the context of our research question.

²² The coefficients for temporary contracts in specifications (4) and (5) are not significantly different from each other. Including the EPL measures as selection instruments has very little effect, which suggests that the country fixed effects absorb most of the explanatory power of the national differences in employment regulations.

²³ Because of the continuous measures of unemployment and EPL, the local ATE is not very informative since it refers to an unidentifiable segment of the population.

Table 4.4: Estimations of on-the-job informal learning

	OLS (1)	2SLS - LATE (2)	2SLS - LATE (3)	FIML - ATE (4)	FIML - ATE (5)
A. Informal Learning					
Temporary contract	0.089*** (0.021)	1.375*** (0.392)	1.019*** (0.312)	0.223*** (0.045)	0.221*** (0.045)
Age	-0.027*** (0.005)	0.016* (0.010)	0.009 (0.009)	-0.023*** (0.004)	-0.023*** (0.004)
Age ² (*100)	0.018*** (0.006)	-0.032*** (0.012)	-0.023** (0.011)	0.017*** (0.005)	0.017*** (0.005)
Years of education	0.019*** (0.003)	0.022*** (0.004)	0.021*** (0.003)	0.019*** (0.003)	0.019*** (0.003)
Overeducated	-0.102*** (0.015)	-0.126*** (0.017)	-0.120*** (0.016)	-0.105*** (0.015)	-0.105*** (0.015)
Undereducated	0.138*** (0.021)	0.137*** (0.024)	0.137*** (0.023)	0.137*** (0.021)	0.137*** (0.021)
Learning attitude (std.)	0.203*** (0.010)	0.203*** (0.010)	0.203*** (0.010)	0.203*** (0.010)	0.203*** (0.010)
Numeracy skills (std.)	-0.033*** (0.008)	-0.015 (0.010)	-0.019** (0.009)	-0.031*** (0.008)	-0.031*** (0.008)
Tenure	-0.003*** (0.001)	0.006* (0.003)	0.005*** (0.002)	-0.002** (0.001)	-0.002** (0.001)
Working hours	0.007*** (0.001)	0.008*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
B. Temporary Contract					
Unemployment rate (std.)		0.048*** (0.011)	0.041*** (0.003)	AME 0.015*** (0.004)	AME 0.011*** (0.004)
EPL permanent (std.)			0.405*** (0.148)		0.308*** (0.116)
EPL temporary (std.)			-0.222** (0.093)		-0.193*** (0.062)
Unemployment*EPL permanent			0.020** (0.008)		0.012** (0.005)
Unemployment*EPL temporary			0.007** -0.003		0.009*** (0.003)
Age		-0.019*** (0.003)	-0.020*** (0.003)	-0.017*** (0.003)	-0.017*** (0.002)
Age ² (*100)		0.024*** (0.003)	0.025*** (0.003)	0.021*** (0.003)	0.021*** (0.003)
Years of education		-0.002 (0.001)	-0.002 (0.001)	0.000 (0.001)	0.000 (0.001)
Overeducated		0.018*** (0.005)	0.017*** (0.005)	0.016*** (0.004)	0.016*** (0.004)
Undereducated		0.001 (0.007)	0.001 (0.007)	0.000 (0.008)	0.000 (0.008)
Learning attitude (std.)		0.001 (0.003)	0.001 (0.003)	0.001 (0.002)	0.001 (0.002)
Numeracy skills (std.)		-0.014*** (0.003)	-0.014*** (0.003)	-0.014*** (0.003)	-0.014*** (0.002)
Tenure		-0.007*** (0.000)	-0.007*** (0.000)	-0.009*** (0.001)	-0.009*** (0.001)
Working hours		-0.001 (0.000)	-0.001 (0.000)	-0.001** (0.000)	-0.001** (0.000)
<i>Other controls</i>		<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
<i>First-stage tests</i>		F(51, 25800) =40.0	F(55, 25796) =34.45	Wald chi2(51) =3610.4	Wald chi2(55) =3636.0
<i>Adj. R² first-stage</i>		0.144	0.145	0.192	0.194
<i>Correlation of errors (ρ)</i>				-0.085*** (0.023)	-0.085*** (0.023)
<i>IV test of endogeneity / FIML Wald test of independent equations (ρ=0)</i>		F(1,215)=61.9 (p=0.000)	F(1,215)=39.2 (p=0.000)	Chi ² (1)=14.1 (p=0.000)	Chi ² (1)=14.0 (p=0.000)
<i>IV overidentification test</i>			Chi ² (4)=7.3 (p=0.121)		
<i>LR test against unconstrained model</i>				Chi ² (1)=25.9 (p=0.000)	Chi ² (5)=51.4 (p=0.000)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The informal learning dependent variable is standardised. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is used as a selection instrument in columns (2) and (4) and columns (3) and (5) add the interaction of the latter with the corresponding standardised EPL indexes for permanent and temporary employment. All the estimations include a constant term. Other controls include dummies for occupation, industry, firm size, year of interview and country-fixed effects. The term AME denotes average marginal effects. N=25,853.

Concerning the admissibility of our instrument, the Wald and F-tests after non-linear and linear first-stage estimations, respectively show that the unemployment rate, in addition to the other covariates, makes a significant contribution to the temporary contract prediction. Furthermore, the LR-tests in the models (4) and (5) confirm that the instrumented models fit the data significantly better than the constrained model that does not include any instrument.

Most of our control variables affect the dependent variable according to the expectations from human capital theory (e.g. Ben-Porath, 1967; Heckman, 1978; Killingsworth, 1982). We find a quadratic relation between on-the-job informal learning and age, suggesting a minimum turning point of investments at the end of employees' working life.²⁴ Years of education are positively correlated with learning in the workplace, although overeducated employees tend to invest less in informal learning than workers in a well-matched job, while undereducated employees invest more.²⁵ There is also a positive relation between informal learning intensity and learning attitude and working hours and a negative relation with tenure and the numeracy skills score. We also find that informal learning tends to be significantly higher for individuals employed in high-skilled occupations and larger firms. As discussed in the introduction, our main finding could be rationalised by several mechanisms. Even though we cannot test all of them with our data, in Sections 4.5.3 and 4.5.4, we present additional results that provide further insights into these possible mechanisms.

4.5.2 Robustness of Main Results

In assessing the robustness of our main findings, the most important concern is the possible misspecification error of the contract selection equation (4.2). The first issue we address is the sensitivity of our main estimations to different specifications of the probit model. We tested a range of models in contrast to our FIML estimates in Table 4.4 and present the results in the Appendix, Table 4.A2.

Specifications in Panel A of Table 4.A2 exclude the variables we included as determinants of temporary contract selection in our main estimations in Table 4.4. The table shows that excluding these regressors only from the selection equation (4.2) slightly lowers the estimated ATE. However, the resulting coefficients are not significantly different from our main estimate, which suggests that our model is robust. Only when country dummies are excluded from the probit model, we observe an increase in the estimated ATE from 0.22 to 0.28. This suggests that country-fixed effects are important controls for the type of contract selection. The increase in the ATE is, however, not excessive, since the unemployment rate seems to capture these country differences in the absence of the country-fixed effects in the probit equation (the average marginal effect of the unemployment rate increases from 0.015 to 0.027).

²⁴ The estimated average minimum (68 years) at which informal learning starts increasing with age is not very meaningful. However, as shown in Section 4.4.1, this is because investments in informal learning differ by age between permanent and temporary employees.

²⁵ Our estimations control for the fact that workers have a job at the appropriate educational level. Nonetheless, estimations that do not control for educational mismatches yield to very similar results (see Table 4.A2).

Specifications in Panel B of Table 4.A4 show how the main estimated coefficient of temporary contracts changes with the gradual introduction of control variables. Panel C shows the results of various alternative specifications. First, we estimate our main model including age as a categorical variable (11 dummies of five-year age groups) instead of the continuous variable. The resulting coefficient is 0.208, with a standard error of 0.046; statistically the same size and significance of our main estimates in Table 4.4. Second, we include a polynomial of the second order for tenure in an alternative specification. Again, the resulting coefficients are very similar to our main results in Table 4.4.

Last, we run several other sensitivity analyses of our main specification, using various restricted samples. The results are shown in Panel D of Table 4.A2. First, we exclude the individuals with a corresponding unemployment rate higher than 20% and 10%, as previously described in Figure 4.1. The resulting estimates are slightly lower but still the estimated ATE of interest remains significant and of similar magnitude as that of Table 4.4. We also estimate our main model excluding non-European countries (i.e. Canada, Japan, and Korea). In this case, the estimate of temporary contracts rises from 0.22 to 0.28 of a standard deviation. This indicates, as expected, some heterogeneity of the effect between countries, and shows the robustness of our main result.

To test for the possibility that our results are driven by a few countries, we continue with other specifications that exclude the countries with the largest shares of temporary employees, that is, Poland, Korea, Spain and Slovak Republic. Similarly, in the next robustness test, we also exclude from our sample those countries with the highest unemployment rates: Poland, Germany and Slovak Republic.

Finally, we test our main model restricting the sample to prime age workers, i.e., employees in the age ranges 20-64, 25-64, and 25-54. In all cases, the estimated ATE of interest remains significant and of similar magnitude as that of Table 4.4. Only the estimate from the sample of workers in ages 25-64 slightly declines from 0.22 to 0.18 of a standard deviation, which is explained later in Section 4.5.4.1 by the decreasing effect of temporary contracts on informal learning with age. This can be verified here by observing the results of the prime age workers (25-54) sample, which gives a significant coefficient of 0.21 of a standard deviation. It is important to note that the predicted values of ρ remain negative in all these robustness specifications presented in Table 4.A2, indicating that our main results hold. The Wald tests are all significant with 95% confidence.

4.5.3 Informal Learning and Training: Substitution or Complementarity?

A plausible explanation for the higher intensity of informal learning of temporary workers could be that on-the-job informal learning substitutes for the lack of formal training. We test for this possibility in this section.

4.5.3.1 Training incidence

In this subsection, we perform estimations to first validate in our sample the negative association of temporary contracts with training participation²⁶ found in various previous studies.

The results in Table 4.5 confirm that the temporary contract indicator in our sample yields the expected negative coefficient. The FIML results in columns (2) and (3) indicate that the probit estimation (1) is biased downwards to some extent. Once the selection into the contract types is controlled for, we find that workers in temporary jobs are, on average, 2.5 percentage points less likely to participate in training activities than those in permanent employment are. The coefficients of the covariates included in these estimations are consistent with human capital theory and findings in previous empirical literature (e.g. Booth, 1991; Green, 1993).

The negative value of ρ suggests that unobservables that decrease temporary contract selection probably occur with unobservables that increase training participation, as suggested in Section 4.3. For the FIML²⁷ specifications (3) and (4), the Wald tests indicate that we can reject the null hypothesis of $\rho=0$ with 95% confidence.

4.5.3.2 Complementarity

The finding that workers with temporary contracts are less likely to participate in training but engage more intensively in informal learning raises the question of whether informal learning is a training substitute for temporary workers.

To answer this question, we first observe whether there is a difference in the informal learning intensity of employees who undertook training and those who did not. Figure 4.2 illustrates this difference among temporary and permanent workers separately. This figure suggests complementarity between training and informal learning, regardless of the type of contract, since the intensity of investments in the latter is shown to be consistently larger when following training. This result holds for both temporary and permanent employees.

To test whether there is, indeed, complementarity between training and informal learning, we include training participation and its interaction with temporary contracts in our main equation for informal learning. To provide reliable results, we use an additional selection instrument to correct for the potential endogeneity of formal training in the informal learning equation. To this end, we

²⁶ Our training variable is a dummy of participation in job-related training during the previous 12 months. It is based on the following questions: During the last 12 months, have you 1) participated in courses conducted through open or distance education, 2) attended any organised sessions for on-the-job training, 3) participated in seminars or workshops, or 4) participated in courses or private lessons not already reported? This variable takes the value one if the individual participated in any of these job-related activities and zero otherwise. The response rate to these questions was 90%, which reduced the sample size for this analysis to 23,232 observations.

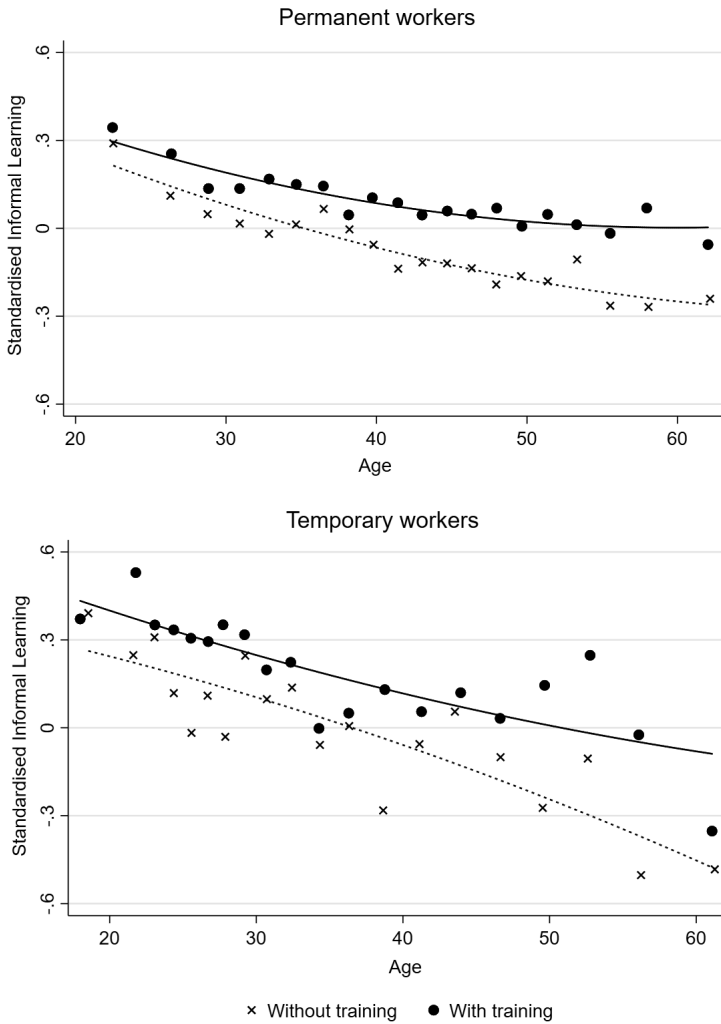
²⁷ Since the endogenous SRM for a binary outcome follow a different structure, the Stata command *etregress* is inappropriate. We therefore used the *gllamm* program to fit the correspondent models (3) and (4). For a detailed description of this command, see Miranda and Rabe-Hesketh (2006); Rabe-Hesketh, Skrondal, and Pickles (2005).

Table 4.5: Estimations of job-related training participation

	Probit (1)	FIML (2)	FIML (3)
A. Training			
Temporary contract	<i>AME</i> -0.054*** (0.010)	<i>AME</i> -0.026*** (0.010)	<i>AME</i> -0.025*** (0.010)
Age	0.010*** (0.002)	0.009*** (0.002)	0.009*** (0.002)
Age ² (*100)	-0.015*** (0.002)	-0.013*** (0.002)	-0.013*** (0.002)
Years of education	0.012*** (0.001)	0.014*** (0.001)	0.014*** (0.001)
Overeducated	-0.017** (0.007)	-0.016*** (0.005)	-0.016*** (0.005)
Undereducated	0.028*** (0.011)	0.025*** (0.008)	0.025*** (0.008)
Learning attitude (std.)	0.035*** (0.004)	0.034*** (0.003)	0.034*** (0.003)
Numeracy skills (std.)	0.033*** (0.004)	0.029*** (0.000)	0.029*** (0.000)
Tenure	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Working hours	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
B. Temporary Contract			
Unemployment rate (std.)		<i>AME</i> 0.014*** (0.003)	<i>AME</i> 0.011*** (0.003)
EPL permanent (std.)			0.382*** (0.130)
EPL temporary (std.)			-0.230*** (0.079)
Unemployment*EPL permanent			0.011** (0.005)
Unemployment*EPL temporary			0.012*** (0.003)
<i>Other controls</i>		<i>yes</i>	<i>yes</i>
<i>First-stage tests</i>		Wald chi ² (51) =3221.3	Wald chi ² (55) = 3248.3
<i>Adj. R2 first-stage</i>		0.192	0.193
<i>Correlation of errors (ρ)</i>		-0.078** (0.035)	-0.082** (0.037)
<i>Wald test of independent equations (ρ=0)</i>		Chi ² (1)=4.12 (p=0.042)	Chi ² (1)=4.05 (p=0.044)
<i>LR test against unconstrained model</i>		LR chi ² (1)=10.2 (p=0.001)	LR chi ² (5)=18.8 (p=0.002)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The training participation dependent variable is binary. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is used as a selection instrument in column (3) and column (4) adds the interaction of the latter with the corresponding standardised EPL indexes for permanent and temporary employment. All the estimations include a constant term and the same control variables as reported in Table 4.4. The term AME denotes average marginal effects. N=23,232.

Figure 4.2: Complementarity between training and informal learning by type of contract



apply the identification approach of Leuven and Oosterbeek (2008), which suggests that having missed out on formal training opportunities for unexpected reasons is a valid approximation to a random assignment to training. This is achieved by using the information obtained through two designed survey questions. The first is whether there was any training related to work that the respondent wanted to attend but did not do so. The second asks whether this non-participation was due to some random event such as family circumstances, transient illness, or any other unexpected personal or job-related incident. As shown in Panel C of Table 4.6, having missed a training activity due to a random or unexpected event indeed increases the probability of training participation by approximately 6 percentage points.

Table 4.6: Estimations of complementarity between training and informal learning

	FIML (1)	FIML (2)	FIML (3)	FIML (4)
A. Informal Learning				
Temporary contract	0.224*** (0.051)	0.221*** (0.050)	0.220*** (0.045)	0.217*** (0.044)
Training	0.190*** (0.014)	0.190*** (0.014)	0.074*** (0.015)	0.074*** (0.015)
Training*Temporary contract	0.001 (0.039)	0.001 (0.039)	-0.009 (0.042)	-0.009 (0.042)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
B. Temporary Contract				
	<i>AME</i>	<i>AME</i>	<i>AME</i>	<i>AME</i>
Unemployment rate (std.)	0.014*** (0.003)	0.011*** (0.003)	0.014*** (0.003)	0.010*** (0.003)
EPL permanent (std.)		0.372*** (0.129)		0.361*** (0.131)
EPL temporary (std.)		-0.232*** (0.074)		-0.221*** (0.076)
Unemployment*EPL permanent		0.011** (0.005)		0.011** (0.005)
Unemployment*EPL temporary		0.011*** (0.003)		0.011*** (0.003)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
C. Training				
			<i>AME</i>	<i>AME</i>
Training missed (random event)			0.064*** (0.023)	0.064*** (0.023)
<i>Other controls</i>			<i>yes</i>	<i>yes</i>
<i>Correlation of errors (ρ)</i>	-0.082*** (0.022)	-0.079*** (0.022)	-0.082*** (0.022)	-0.081*** (0.022)
<i>Wald test of independent equations ($\rho = 0$)</i>	Chi ² (1) = 13.9 ($\rho = 0.000$)	Chi ² (1) = 13.6 ($\rho = 0.000$)	Chi ² (1) = 13.4 ($\rho = 0.000$)	Chi ² (1) = 13.2 ($\rho = 0.000$)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The training participation dependent variable is binary. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is used as a selection instrument in columns (1) and (3), and columns (2) and (4) add the interaction of the latter with the corresponding standardised EPL indexes for permanent and temporary employment. The training missed independent variable is binary. All the estimations include a constant term and the same control variables as reported in Table 4.4. The term AME denotes average marginal effects. N=23,232.

Panel A of Table 4.6 shows that the positive relation between informal learning and job-related training holds after controlling for various individual and employer characteristics. Specifications (3) and (4) that account for the endogeneity of both temporary contracts and training participation show that, on average, participation in training activities increases informal learning by 0.74 of a standard deviation. The magnitude of this complementarity does not differ by type of contract, since the interaction term is close to zero and statistically insignificant. This finding shows that both the direction and size of the complementarity between informal learning and job-related training are equal for temporary and permanent employees, which suggests that the higher informal learning investment of temporary workers does not seem to substitute for their lack of formal training. Moreover, this result indicates that temporary workers engage more intensively in informal learning, even after controlling for training participation.

4.5.4 Heterogeneous Effects in Informal Learning

To provide further insights on the possible explanation of our main results, in this subsection, we analyse two types of heterogeneity: first, the heterogeneous learning investments of temporary and permanent workers that could result from observed individual and job-content characteristics, which allows us to distinguish between ATE and ATT; and, second, heterogeneous effects that could result from unobserved personal characteristics, which enable us to estimate MTEs.

4.5.4.1 Heterogeneous workers

Although temporary workers are, on average, more intensively engaged in informal learning, this could differ among workers with different characteristics. Temporary employees could, for instance, have different expectations for their career prospects. If that is the case, we might expect younger workers and those with shorter tenure to have stronger incentives to engage in informal learning when they are employed in a temporary job, since this might help them to acquire a permanent contract.

To investigate this heterogeneity, we estimate some endogenous SRM of informal learning that allow covariates to vary by contract type, as explained in Section 4.3. The significant interaction results are shown in Table 4.7, which indicate that when allowing for heterogeneous responses to temporary contracts, our main conclusion holds. Both the ATE and ATT remain significant and positive, the latter being of similar size to our estimation in Table 4.4. These heterogeneous models of informal learning show that the coefficients of age, tenure, and working hours²⁸ differ by type of contract, while years of education, educational mismatches, learning attitude and numeracy skills do not. These results confirm our expectations that the rates at which informal learning decreases with

²⁸ Informal learning increases with working hours only in the case of permanent workers. The number of working hours does not seem to affect the informal learning intensity of temporary employees.

age²⁹ and tenure are significantly greater for those with a temporary contract. This finding suggests that our main result is particularly driven by temporary employees who are younger and have lower tenure. More precisely, being a year older and having one additional year of tenure decreases informal learning by 0.026 and 0.008 of a standard deviation, respectively, in the case of temporary workers, whereas, for permanent employees, the decreases are 0.022 and 0.002 of a standard deviation.

As mentioned previously, the larger investments in informal learning of temporary workers are expected to be more beneficial earlier in the working career, when the workers probably have better prospects of gaining a permanent position. This suggests that the difference in informal learning investments between temporary and permanent workers ends gradually and will vanish at some age and after some years of tenure. Nonetheless, it is not just a young, early career effect. According to estimation (2) in Table 4.7, the positive ATE of temporary contracts become insignificant (at the 95% confidence level) after the age of 46. Similarly, the positive ATE of temporary contracts disappear after approximately 8 years of tenure (see Figure 4.3). This could probably be due to workers adjusting their labour mobility expectations when they feel trapped in a temporary job with no career prospects.

If workers generally have stronger preferences for permanent contracts (Booth et al., 2002; Holmlund and Storrie, 2002), it is reasonable to think that flexible workers rationally invest more in on-the-job informal learning to increase their chances of promotion to a more stable/secure job. Thus, we could expect that employees with expectations of upward mobility in the labour market are more likely to invest in informal learning. Since we cannot directly test for this mechanism with the PIAAC data, we use data from the European Skills Survey (2014) that allow us to distinguish between temporary workers with good and poor prospects of job stability, for some additional descriptive analyses.

Estimation results using data from the European Skills Survey,³⁰ and the application of a similar endogenous SRM that accounts for the endogeneity of selection into the type of contract,³¹ indicate that full-time workers with a tem-

²⁹ Table 4.7 also shows that the coefficient of age squared is not significantly different from zero for temporary employees, suggesting that the minimum turning point of learning investments at the end of employees' working lives holds only for workers with a permanent contract.

³⁰ In this survey, the informal learning measure is a categorical variable derived from the following question: How often, if at all, does your job involve learning new things? The respondent's options were never, sometimes, usually, and always. We constructed a temporary contract dummy variable that takes the value one for fixed-term/temporary agency contracts and zero for indefinite/permanent contracts. We derive the dummy variable on the prospects of job stability from the following question: How likely or unlikely do you think it is that you will lose your job in the next two years? Please use a scale of from 0 to 10, where 0 means very unlikely and 10 very likely. This variable takes the value of one (for a good prospect of job stability) if the answer given was below five and zero (indicating poor prospects of job stability) if the answer given was five or above.

³¹ In this application, we also use the corresponding unemployment rate at the age-group-by-country level as the selection instrument. We use data for 2006, collected from the OECD Statistics website, seven years prior to the survey to avoid likely global crisis changes affecting the comparability of these results with our main results derived from the PIAAC data. Since the informal learning information provided in this dataset has an ordered structure, we implemented an endogenous SRM that fits an ordered probit model for the dependent variable. For this purpose, we used the `cmp` program. For a detailed description of this

Table 4.7: Estimations of on-the-job informal learning - heterogeneous employees

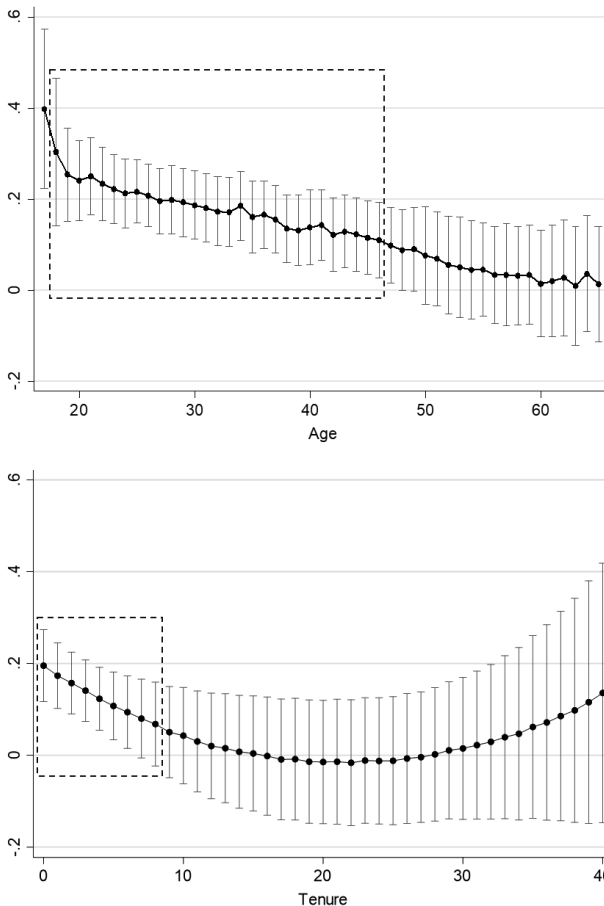
	FIML (1)	FIML (2)
<i>A. Informal Learning</i>		
<i>ATE</i>	0.179*** (0.047)	0.179*** (0.047)
<i>ATT</i>	0.206*** (0.046)	0.206*** (0.046)
Age	-0.022*** (0.004)	-0.022*** (0.004)
Age * TC	-0.004** (0.002)	-0.004** (0.002)
Age ² (*100)	0.018*** (0.005)	0.018*** (0.005)
Age ² (*100) * TC	-0.018*** (0.006)	-0.018*** (0.006)
Tenure	-0.002*** (0.001)	-0.002*** (0.001)
Tenure * TC	-0.006** (0.002)	-0.006** (0.002)
Working hours	0.008*** (0.001)	0.008*** (0.001)
Working hours * TC	-0.007*** (0.002)	-0.007*** (0.002)
<i>Other controls and interactions</i>	<i>yes</i>	<i>yes</i>
<i>B. Temporary Contract</i>		
Unemployment rate (std.)	AME 0.015*** -0.003)	AME 0.012*** (0.003)
EPL permanent (std.)		0.314*** (0.115)
EPL temporary (std.)		-0.194*** (0.063)
Unemployment*EPL permanent		0.012** (0.005)
Unemployment*EPL temporary		0.009*** (0.002)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>
<i>Correlation of errors (ρ)</i>	-0.049*** (0.017)	-0.049*** (0.017)
<i>Wald test of independent equations ($\rho = 0$)</i>	Chi ² (1) = 9.04 (p = 0.002)	Chi ² (1) = 8.52 (p = 0.003)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The informal learning dependent variable is standardised. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is used as a selection instrument in column (1) and column (2) adds the interaction of the latter with the corresponding standardised EPL indexes for permanent and temporary employment. All the estimations include a constant term and the same control variables as reported in Table 4.4. The term AME denotes average marginal effects. N=25,853.

porary contract particularly engage more intensively in informal learning on the job than those with a permanent contract, when the temporary contract offers opportunities for future job stability. Tables 4.8 and 4.9 show the respective results. In these data, we also observe that the likelihood of job stability perceived by employees with a temporary contract decreases with the worker's age and years of tenure with the same employer.

These analyses provide further insights into the theoretical hypothesis (e.g. Weiss, 1986) that expectations of transition to permanent (more stable) employment could stimulate workers to make supplementary investments in informal learning while in a temporary job as one of the mechanisms driving our main findings. We discuss this issue more in detail in Section 4.6.

Figure 4.3: Estimated ATE of temporary contracts on informal learning over age and tenure.



Notes: This figure shows contrasts of Linear Prediction and Marginal effects computed based on the endogenous switching regression model (2) in Table 4.7

command, see Roodman (2011).

Table 4.8: Estimation coefficients of informal learning by type of contract, European Skills Survey 2014

	OLS (1)	OProbit (2)	FIML (3)	OLS (4)	OProbit (5)	FIML (6)
A. Informal Learning						
Temporary contract	0.089*** (0.002)	0.126*** (0.030)	0.239*** (0.064)	0.047* (0.027)	0.064 (0.040)	0.166** (0.067)
Good prospects of job stability				0.015 (0.017)	0.028 (0.025)	0.023 (0.023)
Temporary contract * Good prospects				0.112*** (0.039)	0.167*** (0.055)	0.209*** (0.060)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
B. Temporary Contract						
Unemployment rate (std.)			AME 0.014*** (0.004)			AME 0.014*** (0.004)
<i>Other controls</i>			<i>yes</i>			<i>yes</i>
<i>Correlation of errors (ρ)</i>			-0.065** (0.031)			-0.069** (0.032)
<i>Wald test of independent equations ($\rho = 0$)</i>			Chi ² (1)=4.81 (p=0.028)			Chi ² (1)=4.74 (p=0.029)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. Data come from the European Skills Survey, conducted in 21 European countries in 2014. The sample includes full-time workers aged 24-65 not enrolled in any formal education programme. The informal learning dependent variable is ordered in four categories (0-3). Good prospects of job stability refers to a low self-reported probability of losing the job in the next two years. Other controls include age, age square, educational level, education mismatch, tenure, working hours, learning attitude, occupation, industry, firm size and country dummies. The standardised unemployment rate for 2006 (at the age-group-by-country level, collected from the OECD statistics) is used as a selection instrument in the FIML estimations (3) and (6). The term AME denotes average marginal effects. N=17,442.

Table 4.9: AME of temporary contracts on informal learning, European Skills Survey 2014

	<i>Never</i>	<i>Sometimes</i>	<i>Usually</i>	<i>Always</i>
<i>Temporary contract (Perm. contract ref.)</i>				
With good prospects of job stability	-0.007** (0.003)	-0.048** (0.021)	0.016*** (0.006)	0.039** (0.018)
With poor prospects of job stability	-0.015*** (0.002)	-0.123*** (0.023)	0.028*** (0.002)	0.110*** (0.023)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. This table shows average marginal effects (AMEs) computed on the FIML specification (6) of Table 4.8. The dependent variable informal learning is measured by four ordinal categories as shown in the four columns. The AME for categorical variables is the discrete change from the base level. N=17,442.

4.5.4.2 Heterogeneous jobs

Our main estimates could also differ if firms select workers into temporary versus permanent contracts because of the task content of different jobs, which could lead to heterogeneity in informal learning opportunities. For instance, it could be that high-skilled jobs or jobs that entail more task flexibility and problem-solving tasks are less likely to be selected for temporary contracts³² and, at the same time, offer more informal learning opportunities. To test for this heterogeneity, we include different job-content characteristics and estimate heterogeneous effects in our interacted endogenous SRM.

First, we do not find any evidence that informal learning intensity differs between high-skilled and low-skilled temporary jobs, although those employed in high-skilled jobs are significantly less likely to be selected into temporary contracts and simultaneously tend to engage more intensively in learning on the job.³³ Second, by using the information on task-content of jobs as provided by the PIAAC survey, we are able to test for further heterogeneity.

Table 4.10 shows that jobs that involve greater task flexibility, and more planning, literacy-related tasks and ICT-related tasks are less likely to be filled with individuals employed on a temporary basis. The contrary occurs with jobs that more often involve physical tasks, while task collaboration and teamwork and problem-solving and numeracy-related tasks do not seem to have a significant influence in job contract type selection. At the same time, Table 4.10 shows that employees tend to engage more intensively in informal learning when they are employed in jobs with higher levels of task flexibility and teamwork and problem-solving, planning, numeracy-, literacy- and ICT-related tasks. Conversely, jobs with more physical-task content seem to offer fewer informal learning opportunities. More interestingly, the interaction coefficients suggest no significant differences between temporary and permanent workers regarding the above results, the only exception being the larger positive effect on informal learning that

³² Descriptive statistics are provided in Table 4.A3 in the Appendix.

³³ The interaction term between temporary contracts and high-skilled occupations is not significant with 90% of confidence. Results are available upon request.

task collaboration and teamwork has for temporary employees, which suggests that interaction with peers particularly enhances on-the-job informal learning of temporary workers. When including all these tasks characteristics, we still find the positive influence of temporary contracts on informal learning.

4.5.4.3 Heterogeneous effects from unobservable characteristics

There are reasons to expect further unobservable heterogeneity in the informal learning outcomes of employees with a temporary contract. Since workers differ, for instance, in their ability, personality traits, and the information and expectations they have about their future career when starting a job, individuals with different unobservable characteristics could respond differently to a temporary contract. Heterogeneity could also result from variety in the quality of temporary contracts, for example, regarding the chances of these being converted into permanent contracts.

To investigate this heterogeneity, we estimate MTEs (at the mean value of all covariates in X) along the 99 percentile points of the distribution of unobservables U_D , computed on an endogenous SRM of informal learning similar to specification (5) of Table 4.4 but allowing different correlation parameters ρ and variance σ^2 between permanent and temporary workers, as explained in Section 4.3. The corresponding results are shown in Figure 4.4 and Table 4.A4 in the Appendix. When allowing for this heterogeneity, our main findings hold. The model fits well and, in fact, the estimated ATE increases from 0.22 to 0.28 of a standard deviation in comparison with our results in Table 4.4.

The MTE estimates describe whether informal learning outcomes for workers on the margin of a temporary job placement increase or decrease with the probability of being selected into a temporary contract, $p_i(Z_i)$, conditioned on marginal changes in the instrument. The MTEs then show if and to what extent the investments in informal learning of employees change if they are shifted into a temporary contract by marginally increasing the corresponding unemployment rate.

Following Heckman and Vytlacil (2005), in Figure 4.4 the MTEs of low U_D values show the expected effects in the informal learning of workers who are more likely, due to unobserved reasons, to be selected into temporary contracts, that is, those who would be selected even if the unemployment rate was rather small. High U_D values in Figure 4.4 therefore represent workers with higher probabilities of being selected into a permanent contract, which means that the unemployment rate would have to be very large to induce them to be in a temporary job. The upward-sloping shape of the MTE curve in U_D in Figure 4.4 then indicates that the ATE of temporary contracts is higher for employees who are more likely to be selected in permanent jobs (high U_D), which means that the lower the likelihood of selection into a temporary contract, the higher the informal learning investments of workers are expected to be if they were in a temporary rather than in a permanent job.³⁴ This increase in the MTEs at

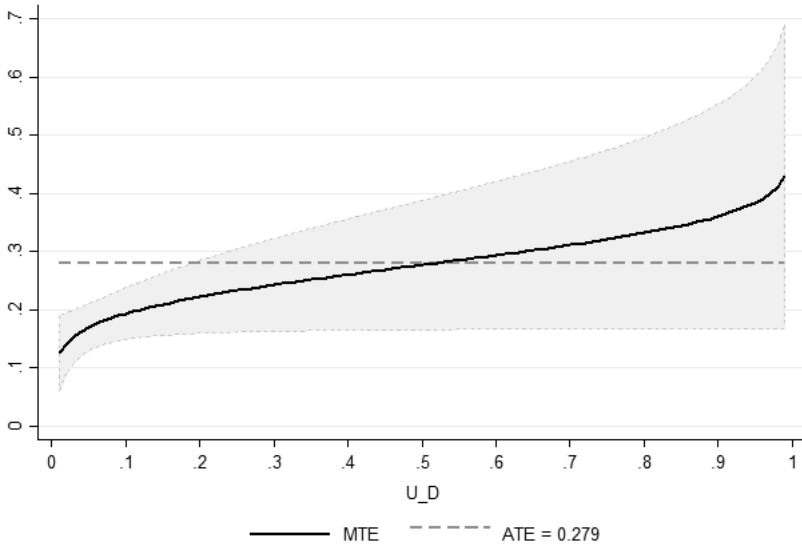
³⁴ The increasing MTE curve also indicates negative selection in unobserved characteristics, in line with our hypotheses in Section 3. This negative selection is also suggested if $\rho_1 > \rho_0$ Cornelissen, Dustmann, Raute, and Schönberg (2016), as shown in Table 4.A4 in the Appendix.

Table 4.10: Heterogeneous effects by job content characteristics

	FIML (1)	FIML (2)
A. Informal Learning		
<i>ATE</i>	0.245*** (0.044)	0.244*** (0.045)
<i>ATT</i>	0.261*** (0.045)	0.260*** (0.046)
Task flexibility	0.015** (0.007)	0.015** (0.007)
Task flexibility * TC	-0.001 (0.022)	-0.001 (0.022)
Task collaboration - teamwork	0.166*** (0.009)	0.166*** (0.009)
Task collaboration - teamwork * TC	0.049** (0.020)	0.049** (0.020)
Problem-solving tasks	0.136*** (0.008)	0.136*** (0.008)
Problem-solving tasks * TC	0.004 (0.020)	0.004 (0.020)
Planning tasks	0.029*** (0.007)	0.029*** (0.007)
Planning tasks * TC	-0.025 (0.020)	-0.025 (0.020)
Physical tasks	0.007 (0.008)	0.007 (0.008)
Physical tasks * TC	-0.012 (0.021)	-0.012 (0.021)
Numeracy-related tasks	0.031*** (0.010)	0.031*** (0.010)
Numeracy-related tasks * TC	-0.003 (0.025)	-0.003 (0.025)
Literacy-related tasks	0.151*** (0.011)	0.151*** (0.011)
Literacy-related tasks * TC	-0.034 (0.032)	-0.034 (0.032)
ICT-related tasks	0.048*** (0.012)	0.048*** (0.012)
ICT-related tasks * TC	0.006 (0.030)	0.006 (0.030)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>
B. Temporary Contract		
	<i>AME</i>	<i>AME</i>
Unemployment rate (std.)	0.015*** (0.003)	0.011*** (0.003)
EPL permanent (std.)		0.249** (0.012)
EPL temporary (std.)		-0.160** (0.071)
Unemployment*EPL permanent		0.011** (0.005)
Unemployment*EPL temporary		0.007** (0.003)
Task flexibility	-0.005** (0.002)	-0.005** (0.002)
Task collaboration and teamwork	0.003 (0.002)	0.003 (0.002)
Problem-solving tasks	0.001 (0.003)	0.001 (0.003)
Planning tasks	-0.007*** (0.002)	-0.007*** (0.002)
Physical tasks	0.009*** (0.003)	0.009*** (0.003)
Numeracy-related tasks	0.000 (0.003)	0.000 (0.003)
Literacy-related tasks	-0.009** (0.004)	-0.009** (0.004)
ICT-related tasks	-0.015*** (0.004)	-0.015*** (0.004)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>
<i>Correlation of errors (ρ)</i>	-0.095*** (0.027)	-0.094*** (0.027)
<i>Wald test of independent equations ($\rho = 0$)</i>	Chi ² (1) = 12.5 (p = 0.000)	Chi ² (1) = 12.1 (p = 0.001)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The informal learning dependent variable is standardised. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is used as a selection instrument in column (1) and column (2) adds the interaction of the latter with the corresponding standardised EPL indexes for permanent and temporary employment. All the estimations include a constant term and the same control variables as reported in Table 4.4. The term AME denotes average marginal effects. N=23,069.

Figure 4.4: Estimated MTE of temporary contracts on informal.



the upper levels of U_D also suggests that workers on the margin of a temporary job under the highest unemployment rates are likely to have better unobserved characteristics (such as ability and motivation) and thus would invest the most in learning on the job in comparison with their counterparts in permanent jobs. This result could be because these individuals know their odds of obtaining a more stable position, given their own characteristics, and could be more able to make informed choices of investing (or signalling) in learning to improve those possibilities. These MTE results provide more insight into the main mechanism we propose for our findings, that is, that employees in temporary contracts who have better prospects for their future career would rationally invest more in on-the-job informal learning to increase their chances of promotion to a permanent job.

4.6 Conclusions and discussion

We have analysed the difference in informal learning at work between temporary and permanent male employees across 20 OECD countries. Human capital theory predicts that both firms and employees are less willing to invest in skills if workers are hired under temporary contracts. Remarkably, we find that workers in temporary jobs engage more intensively in informal learning than permanent employees do, although the former are, indeed, less likely to participate in formal training.

We conclude that the difference in the intensity of on-the-job informal learning between workers with temporary and permanent contracts is positive and

between 0.22 and 0.28 of a standard deviation.³⁵ These results account for the endogeneity of the selection into temporary contracts by estimating FIML endogenous switching regression models that exploit workers' differential exposure to unemployment and employment protection legislation across countries and age groups.

Our findings could be driven by several mechanisms. Although the PIAAC data do not allow us to assess the particular mechanism behind our main results, we can discard some of the possible explanations and provide further insights on these relevant mechanisms.

First, the hypothesis that temporary workers could substitute informal learning for the lack of formal training is not supported by our analysis. Instead, we find complementarity between these two types of learning, which is consistent with the idea that human capital accumulated through training not only provides workers with higher skills but could also increase their informal learning capacity (Rosen, 1972).

Second, we do not find support for the view that high-skilled workers drive the difference in informal learning between temporary and permanent male employees. Although we find that particularly peer interaction enhances the intensity of informal learning among temporary employees, their difference in informal learning in comparison with permanent workers is not driven by heterogeneity in the task content of temporary jobs.

However, the higher informal learning investments of temporary employees might be driven by the incentives of finding a more stable job. If workers generally prefer permanent contracts (Booth et al., 2002; Holmlund and Storrie, 2002), it is reasonable to think that those in temporary jobs could have stronger incentives to invest in on-the-job learning to increase their chances of promotion to a more secure job.³⁶ In line with this hypothesis, we find that the positive influence of temporary contracts is larger earlier in workers' careers, when workers generally have better prospects of transition to permanent employment than later in their career. Additional descriptive results from the European Skills Survey that allow us to distinguish between temporary workers with and without prospects of job stability in the near future, suggest that the intensity of informal learning is likely higher for temporary workers who have better expectations of job stability.

These results are consistent with our MTE analysis when using the PIAAC dataset, which shows that the expected effects in informal learning are significantly higher among workers with a temporary contract who are more likely to be selected into permanent contracts due to their unobservable characteristics. This finding suggests that those who know about their better chances of obtaining a permanent job position, given their own characteristics, are those expected to show a higher intensity of informal learning at work, which could help to improve these chances in their favour. This analysis suggests that workers may perceive more intense learning as a profitable investment in their career

³⁵ This difference is substantial if we consider that having a temporary contract has a similar impact on informal learning as approximately 10 years of schooling.

³⁶ This could be incentivised by the lower opportunity costs of informal learning in contrast to training (Destré et al., 2008).

development, particularly when facing the uncertainty of a temporary contract. As explained by Weiss (1986), the possibility of shifting to a better job affects the returns to human capital and, therefore, increases the incentives for supplementary investments in learning activities while working. Furthermore, if these investments are positively affected by a lower discount rate because future perspectives become more important, incentives for self-investment increase and give rise to human capital accumulation until a more stable job is obtained.

Our findings could then be likely driven by the mechanism suggested in the literature on the stepping-stone effects of temporary employment. Most of these studies evoke the idea that the odds of transition from temporary to permanent jobs likely increase with investments in on-the-job learning and the improvement of skills while gaining work experience (e.g. Autor, 2001; Booth et al., 2002; Cockx and Picchio, 2012; De Graaf-Zijl et al., 2011; Gagliarducci, 2005). This implies that, if human capital investments on the job decline over the lifecycle by a search for a better job, accepting a temporary job that might pay less initially but involves higher learning potential can be a good strategy for workers in their early careers, since such jobs are more likely to have a stepping-stone effect (Heckman et al., 2002; Sicherman and Galor, 1990).

The implicit stepping-stone incentive of temporary jobs as a plausible explanation for our results also relates to the literature on job matching and screening. If firms use temporary contracts to select the best workers for permanent positions, human capital acquired through learning on the job becomes an important source for firms to investigate the quality of the match and productivity of potential permanent employees (Autor, 2001; Faccini, 2014; Nagypál, 2007). Temporary contracts could then also increase workers' incentives to signal positively their skills and productivity by accumulating more job-specific expertise through on-the-job learning.

Our analyses point towards a potential positive feature of temporary contracts that has important implications. Temporary jobs need not be dead-end jobs. If temporary jobs are taken by individuals in lieu of unemployment in search for further individual promotion in the labour market, these jobs could offer them opportunities for learning by doing particular tasks and productive interaction with other workers. This is important not only as a source of productive accumulation of human capital while working but also as a potential advantage for individuals who would otherwise be unemployed. Such jobs with high learning content might then be a stepping stone towards more stable employment.

The challenge of harnessing the learning potential of temporary employment also implies the reduction of the training participation penalty induced by the short duration of contracts. This penalty might not only contribute to disparities in the labour market but, as we show, could also compromise the benefits of the complementarity between formal training and informal learning. Our complementarity analysis suggests two different kinds of temporary jobs in terms of their learning opportunities: 1) good temporary jobs, with high levels of task autonomy and collaboration, more teamwork and problem-solving tasks, offering good opportunities for training and informal learning and, likely, leading to positive career expectations of upward mobility, and 2) bad temporary jobs,

which have no or very few opportunities to foster workers' human capital. In the latter jobs, workers could get trapped in a cycle between precarious jobs and unemployment. Thus, our study suggests that labour segmentation occurs not only between permanent and temporary jobs but also within temporary employment due to the distinction between temporary jobs of low and high learning content.

Policy makers have already stressed the importance of finding 'an appropriate balance between flexibility and security' to prevent part of the labour force from becoming trapped in dead-end jobs, so-called *flexicurity* agenda (European Commission, 2007). Access to opportunities for workers' human capital development remains a crucial issue for many governments to create such a balance. These policies should therefore be supported by analysing how contract incentives influence workers' skill investments and their career development expectations. The design of these policies underlines the importance of improving firms' learning strategies to optimise the benefits of both training and informal learning as a means of fostering not only successful transitions and the sustainable employability of a flexible workforce but also the aggregate productive capacity of the economy in the long term. In this respect, it is important that further longitudinal research would aim to identify the causal effects of both formal training and informal learning on workers' transition from temporary to permanent jobs.

Appendices to Chapter 4

Table 4.A1: FIML estimations of on-the-job informal learning intensity

	Job-related training (1)	Job-related training (2)	Any type of training (3)	Any type of training (4)
	<i>AME</i>	<i>AME</i>	<i>AME</i>	<i>AME</i>
Unemployment rate t	-0.002 (0.016)	0.009 (0.018)	0.003 (0.017)	0.010 (0.018)
Years of education		0.015*** (0.003)		0.016*** (0.003)
Learning attitude		0.033*** (0.008)		0.033*** (0.008)
Numeracy skills		0.001** (0.000)		0.001*** (0.000)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
<i>N</i>	3,209	3,139	3,348	3,264

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows probit estimation results. The dependent variable is the dummy of participation in training. Unemployment rate t is the standardised unemployment rate (at the age-group-by-country level, same year of the interview) is the independent variable. Other controls include age and country-fixed effects. The term AME denotes average marginal effects.

Table 4.A2: Estimations of informal learning intensity (sensitivity tests)

	ATE (1)	ρ	Wald test $\rho=0$ <i>Chi</i> ² (1)	ATE (2)	ρ	Wald test $\rho=0$ <i>Chi</i> ² (1)
<i>A. Probit specifications excluding</i>						
Overeducated and undereducated	0.179*** (0.045)	-0.062*** (0.023)	7.33 (p = 0.007)	0.177*** (0.045)	-0.061*** (0.023)	7.11 (p = 0.008)
Learning attitude	0.181*** (0.045)	-0.062*** (0.023)	7.3 (p = 0.007)	0.181*** (0.045)	-0.061*** (0.023)	7.08 (p = 0.008)
Numeracy skills	0.184*** (0.047)	-0.066*** (0.024)	7.37 (p = 0.007)	0.181*** (0.047)	-0.065*** (0.024)	7.2 (p = 0.007)
Tenure	0.189*** (0.045)	-0.053*** (0.020)	6.91 (p = 0.009)	0.187*** (0.045)	-0.053*** (0.020)	6.82 (p = 0.009)
Working hours	0.175*** (0.045)	-0.059*** (0.023)	6.35 (p = 0.012)	0.172*** (0.045)	-0.057*** (0.023)	6.08 (p = 0.014)
Occupation	0.179*** (0.045)	-0.059*** (0.023)	6.25 (p = 0.012)	0.176*** (0.045)	-0.057*** (0.023)	5.81 (p = 0.016)
Industry	0.174*** (0.045)	-0.060*** (0.023)	6.75 (p = 0.009)	0.176*** (0.045)	-0.061*** (0.023)	6.9 (p = 0.009)
Firm size	0.182*** (0.045)	-0.062*** (0.022)	7.52 (p = 0.006)	0.180*** (0.045)	-0.062*** (0.022)	7.71 (p = 0.006)
Occupation, industry and firm size	0.178*** (0.046)	-0.062*** (0.023)	7.11 (p = 0.008)	0.175*** (0.046)	-0.061*** (0.023)	6.82 (p = 0.009)
Country fixed effects	0.282*** (0.049)	-0.122*** (0.025)	23.5 (p = 0.000)	0.284*** (0.048)	-0.124*** (0.025)	25.1 (p = 0.000)
<i>B. Adding controls</i>						
Year of interview and country N=26,495	0.394*** (0.058)	-0.168*** (0.028)	36.5 (p = 0.000)	0.393*** (0.045)	-0.167*** (0.028)	35.6 (p = 0.000)
Age, age2 and years of education N=26,338	0.145*** (0.051)	-0.044** (0.021)	4.38 (p = 0.036)	0.143*** (0.050)	-0.042** (0.021)	4.31 (p = 0.038)
Learning attitude and numeracy skills N=26,315	0.154*** (0.044)	-0.052*** (0.020)	6.78 (p = 0.010)	0.151*** (0.044)	-0.051** (0.020)	6.46 (p = 0.011)
Overeducated and undereducated N=26,315	0.158*** (0.044)	-0.050** (0.020)	6.06 (p = 0.014)	0.156*** (0.044)	-0.049** (0.020)	5.8 (p = 0.016)
Occupation N=25,966	0.176*** (0.041)	-0.049** (0.020)	6.33 (p = 0.012)	0.174*** (0.041)	-0.048** (0.020)	6.25 (p = 0.012)
Industry and firm size N=25,853	0.192*** (0.043)	-0.053*** (0.020)	6.85 (p = 0.009)	0.190*** (0.043)	-0.052*** (0.020)	6.65 (p = 0.010)
Tenure and working hours N=25,853	0.223*** (0.045)	-0.085*** (0.023)	14.1 (p = 0.000)	0.221*** (0.045)	-0.085*** (0.023)	14 (p = 0.000)
<i>C. Alternative specifications</i>						
Age variable in 5-year categories N=25,853	0.208*** (0.046)	-0.074*** (0.023)	9.9 (p = 0.002)	0.206*** (0.046)	-0.073*** (0.023)	9.7 (p = 0.002)
Including tenure square term N=25,853	0.212*** (0.040)	-0.101*** (0.020)	23.8 (p = 0.000)	0.213*** (0.040)	-0.101*** (0.020)	24.2 (p = 0.000)
<i>D. Restricted samples</i>						
Unemployment rate <= 0.2 N=25,249	0.190*** (0.045)	-0.066*** (0.023)	7.16 (p = 0.007)	0.188*** (0.045)	-0.066*** (0.023)	6.96 (p = 0.008)
Unemployment rate <= 0.1 N=20,986	0.178*** (0.048)	-0.054** (0.024)	5.14 (p = 0.023)	0.175*** (0.048)	-0.054** (0.024)	4.98 (p = 0.026)
Training sample N=22,232	0.217*** (0.044)	-0.082*** (0.022)	13.5 (p = 0.000)	0.214*** (0.044)	-0.081*** (0.022)	13.1 (p = 0.000)
Excl. non-EU countries N=21,989	0.280*** (0.042)	-0.096*** (0.022)	19 (p = 0.000)	0.279*** (0.042)	-0.096*** (0.022)	18.7 (p = 0.000)
Excl. Poland and Korea N=23,153	0.238*** (0.049)	-0.084*** (0.025)	10.8 (p = 0.001)	0.237*** (0.049)	-0.084*** (0.025)	10.8 (p = 0.001)
Excl. Poland, Korea, Spain, Slovak Rep. N=20,864	0.246*** (0.054)	-0.093*** (0.028)	8.34 (p = 0.004)	0.248*** (0.054)	-0.094*** (0.028)	9.04 (p = 0.003)
Excl. Poland, Germany, and Slovak Rep. N=21,763	0.204*** (0.054)	-0.078*** (0.028)	7.73 (p = 0.005)	0.203*** (0.054)	-0.077*** (0.028)	7.61 (p = 0.006)
Employees in age 20 - 64 N=25,618	0.215*** (0.047)	-0.079*** (0.024)	11.2 (p = 0.000)	0.213*** (0.048)	-0.078*** (0.024)	10.8 (p = 0.001)
Employees in age 25 - 64 N=23,780	0.176*** (0.059)	-0.043** (0.021)	5.27 (p = 0.022)	0.171*** (0.046)	-0.042** (0.021)	4.9 (p = 0.027)
Employees in age 25 - 54 N=19,791	0.210*** (0.042)	-0.075*** (0.023)	10.4 (p = 0.001)	0.204*** (0.041)	-0.071*** (0.023)	9.52 (p = 0.002)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at the age-group-by-country level. The informal learning dependent variable is standardised. The standardised unemployment rate (at the age-group-by-country level, six years prior to the interview) is used as a selection instrument in the estimations presented in column (1), whereas estimations in column (2) add the interaction of the latter with the corresponding standardised EPL indexes for permanent and temporary employment.

Table 4.A3: Other summary statistics

	Permanent		Temporary		Diff. p-value	All	
	Mean	Std. Dev.	Mean	Std. Dev.		Min	Max
<i>Job-content characteristics</i>							
High-skilled occupations	0.47	0.50	0.30	0.46	0.000	0	1
Task discretion (standardised index)*	0.11	0.99	-0.26	1.08	0.000	-2.81	2.43
Task collaboration and team work (standardised index)*	0.06	0.92	0.09	0.99	0.096	-2.69	1.05
Problem-solving tasks (standardised index) *	0.12	0.94	-0.12	1.06	0.000	-2.28	1.17
Planning tasks (standardised index)*	0.06	1.02	-0.14	0.95	0.000	-2.74	3.82
Physical tasks (standardised)	0.01	0.98	0.35	0.95	0.000	-1.09	1.13
Numeracy-related tasks (standardised index)	0.16	0.98	-0.18	1.05	0.000	-1.68	3.55
Literacy-related tasks (standardised index)	0.11	0.94	-0.33	1.17	0.000	-2.82	4.29
ICT-related tasks (standardised index)	0.07	1.01	-0.37	1.04	0.000	-1.37	2.71
<i>Other characteristics</i>							
Job satisfaction	0.01	0.98	-0.19	1.04	0.000	-3.63	1.18
Learning motivation	-0.12	0.98	-0.06	1.07	0.085	-3.32	1.24
Political efficacy perception*	-0.01	1.01	-0.10	1.00	0.000	-1.38	1.83
Social trust*	-0.03	0.97	-0.22	0.88	0.000	-1.41	2.60
Health status	-0.01	0.98	-0.02	1.01	0.86	-1.50	2.59
Observations	22,795		3,058			25,853	

Notes: *Due to lower response rate, we have fewer observations for these variables.

Table 4.A4: MTEs estimates and standard errors

U_D 1	0.128** (0.053)	U_D 30	0.241*** (0.078)	U_D 60	0.306*** (0.102)	U_D 90	0.359** (0.151)
U_D 10	0.192*** (0.044)	U_D 40	0.269*** (0.086)	U_D 70	0.327*** (0.115)	U_D 99	0.437** (0.216)
U_D 20	0.220*** (0.056)	U_D 50	0.282*** (0.098)	U_D 80	0.342** (0.126)		
ATE	0.279*** (0.096)	rho1	-0.112** (0.048)	rho0	-0.054** (0.023)	rho1 - rho0	0.058** (0.026)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors bootstrapped with 200 replications in parentheses (each replication is a bootstrap sample of clusters at the age-group-by-country level). This table shows the average treatment effects (ATEs) for each percentile of the distribution of U_D , computed at the mean value of all covariates in a FIML specification similar to specification (5) of Table 4.4 that allows for different correlation parameters $\rho_1 \neq \rho_0$. N=25,853.

**Work-related learning and skill development in Europe:
Does initial skill mismatch matter?**

Abstract

This chapter provides more insight into the assumption of human capital theory that the productivity of job-related training is driven by the improvement of workers' skills. We analyse the extent to which training and informal learning on the job are related to employee skill development and consider the heterogeneity of this relationship with respect to workers' skill mismatch at job entry. Using data from the 2014 European Skills and Jobs Survey, we find – as assumed by human capital theory – that employees who participated in training or informal learning show greater improvement of their skills than those who did not. The contribution of informal learning to employee skill development appears to be larger than that of training participation. Nevertheless, both forms of learning are shown to be complementary. This complementarity between training and informal learning is related to a significant additional improvement of workers' skills. The skill development of workers who were initially underskilled for their job seems to benefit the most from both training and informal learning, whereas the skill development of those who were initially overskilled benefits the least. Work-related learning investments in the latter group seem to be more functional in offsetting skill depreciation than in fostering skill accumulation.

JEL Classification: J24; M53.

This chapter is joint work with Annemarie Künn-Nelen and Andries de Grip. It has been published with the same title in: *Research in Labor Economics* 45 (2017), p. 345-408. We gratefully acknowledge comments from Jim Allen, Sara de la Rica, Rolf van der Velden and participants of the Cedefop/IZA Workshop on Skills and Skill Mismatch (2015); the Workshop on Education, Skills, and Labor Market Outcomes (2016); and the conferences of the European Society for Population Economics (2016) and European Association of Labour Economics (2016). We thank two anonymous referees of the *Research in Labor Economics* series for their insightful comments, which helped us to significantly improve the paper.

5.1 Introduction

To deal with the challenges of rising global competition, the European Union has set itself goals with respect to formal training and informal learning in the workplace to 'acquire and develop new skills throughout the lifetime of individuals' and increase their productivity (European Commission, 2010b, p. 16). This idea that lifelong learning improves workers' skills builds on human capital theory (Becker, 1964; Ben-Porath, 1967; Heckman, 1976; Mincer, 1962). Several empirical studies have shown that job-related training is associated with higher wages and productivity (e.g. Acemoglu and Pischke, 1999b; Blundell, Dearden, Meghir, and Sianesi, 1999; Görlitz, 2011; Leuven and Oosterbeek, 2008). However, due to lack of data on skill development, there is hardly any empirical literature on the contribution of different forms of human capital investments to workers' skill development.

In this chapter, we provide more insights into the relevance of the assumption that the productivity of training is driven by the improvement of workers' skills. We analyse the extent to which work-related learning is related to the skill development of workers in 28 European countries. We thereby distinguish between formal training participation and informal learning on the job and examine whether the substitution or complementarity between these two types of learning plays a role in workers' skill accumulation. To better understand these relationships, we allow for heterogeneity with regard to employee skill mismatch at job entry.¹

The 2014 European Skills and Jobs Survey shows that, at the start of a job, a significant proportion of the labour force in Europe has skills that either exceed the demands of their job or are insufficient for adequate job performance: 24 percent of all workers report that some of their skills were initially lower than what was required in their job, whereas 25 percent report that their skills were initially higher than required. In comparison with well-matched workers, those who are underskilled, on the one hand, probably need more training or informal learning on the job to perform at an adequate level. We therefore expect job-related learning to add more to these workers' skill level. On the other hand, workers who are overskilled could be more likely to have other reasons to engage in job-related learning, such as keeping their skills up-to-date, which might not reveal skill improvement as such. Because of these differences in the degree of workers' skill utilisation on the job and in the underlying reasons for participating in job-related learning, the skill development of initially mismatched workers is expected to differ from that of workers whose skills fully matched the skill demands in their job.

For this study, we use data on more than 37,000 employees from the European Skills and Jobs Survey, conducted in 2014 by the European Centre for the Development of Vocational Training (Cedefop). This survey is one of the first in which different types of job-related learning as well as employees' skill de-

¹ Since workers' skill mismatch could improve when they participate in learning, we use information on individuals' skill mismatch at the start of the job with their current employer rather than their current skill mismatch status. Hereafter, we use the terms *underskilled*, *well matched*, and *overskilled* to refer to the skill mismatch status of employees at job entry.

velopment and mismatch are measured.² Although these data are not sufficient to establish causality, we provide relevant empirical evidence on the expected relationship between different forms of workplace learning and employee skill accumulation, which has, until now, been a black box in the empirical human capital literature that has mainly focused on workers' wages (De Grip and Sauer-*mann*, 2013).

Our study contributes to the literature in three more ways. First, whereas most empirical studies focus on training participation, we are able to differentiate between formal training and informal learning to investigate in more detail the extent to which workers' participation in different forms of work-related learning contributes to the improvement of their skill levels. Second, we provide more insights into the complementarity between training and informal learning on the job by analysing their interaction in the development of workers' skills. Third, we examine the heterogeneity of the relationships between training and informal learning and workers' skill development with respect to the initial job-worker mismatch status.

In line with the hypotheses derived from skills production models, we find that employees who participate in training or who are more often involved in informal learning show greater skill improvement. The relationship between informal learning on the job and employees' skill development appears to be stronger than that between training participation and skill development. Nonetheless, there is complementarity between these two forms of work-related learning. This complementarity seems to a significant additional improvement of workers' skills. This finding is consistent with the notions of the complementarity and cross-productivity of human capital (Cunha and Heckman, 2007) in an on-the-job context.

We also find that skill mismatches induce heterogeneities and moderate the previous outcomes. Thereby, we find that well-matched, underskilled, and over-skilled employees differ in the extent to which investments in training and informal learning – and their complementarity – are associated with the accumulation of their skills during their working life. A stronger relationship between work-related learning and skill development suggests that initially underskilled employees benefit the most from both training and informal learning, whereas overskilled employees benefit the least. Human capital investments in the latter group seem to be more functional in counteracting skill depreciation and maintaining their skill level than in fostering skill accumulation, as suggested by literature on skill obsolescence (e.g. De Grip and Van Loo, 2002). Our study then suggests that being employed in a skill-challenging job or a job that under-utilises a worker's skills has important implications on the returns to investment in both training and informal learning.

We also analyse the contribution of different types of training and informal learning to workers' skill development. We find that, among well-matched and underskilled employees, training during working hours and training paid by the employer are more strongly related to workers' skill development than training outside working hours and training paid by the employee. Among overskilled workers, however, these differences are rather small and statistically

² For further details on the survey, see Cedefop (2015).

insignificant. With respect to informal learning, we find that, for workers in well-matching jobs, informal learning from colleagues and supervisors seem to contribute to their skills as much as informal learning by trial and error and self-study. Underskilled workers, however, show greater skill improvement when learning by self-study. In contrast, overskilled workers seem to benefit more from informal learning by trial and error and learning from colleagues and supervisors than from self-study, which appears to be insignificant for their skill improvement.

The remainder of the chapter is structured as follows. Section 5.2 discusses the relevant literature. Section 5.3 describes the dataset and the definitions of skill development and skill mismatch as well as the other variables used in the analyses. Section 5.4 describes the estimation method we use – ordered probit models with interaction effects – and explains how to interpret the results. The results are presented in Section 5.5. Section 5.6 concludes the chapter.

5.2 Related Literature

5.2.1 Human Capital Investments and Skill Development

Human capital theory considers on-the-job learning an investment that increases workers' productivity and wages via the accumulation of skills (Becker, 1964; Ben-Porath, 1967; Heckman, 1976; Mincer, 1962). First, at the individual level, most of the empirical literature deals with the relation between training and wages, since hard measures of individual productivity are rare (Acemoglu and Pischke, 1999a; Blundell et al., 1999; Görlitz, 2011; Leuven and Oosterbeek, 2008; O'Connell and Byrne, 2010). An exception is the study of De Grip and Sauermann (2012), who assess the effects of job-related training on individual performance by means of a field experiment. Second, at the firm level, most studies focus on the relation between average training participation and firm productivity as measured by value added (Barrett and O'Connell, 2001; Bartel, 1994; Dearden, Reed, and Van Reenen, 2006; Loewenstein and Spletzer, 1998; Sepúlveda, 2009). However, due to lack of data, the empirical question whether training and informal learning affect workers' performance by increasing workers' skills or whether the performance increase is attributable to other factors still remains (De Grip and Sauermann, 2013).

There is one exception. Green, Ashton, and Felstead (2001) analyse training on and off the job as a determinant of skills supply. Using the British Skills Survey, the authors find that off-the-job training is a determinant of all types of skills included in their analysis except team working, whereas on-the-job training contributes only to workers' problem-solving and team-working skills. However, Green et al. (2001) measure tasks rather than skills by using information on the importance of workers' particular job activities as the dependent variable. Furthermore, their skills measure refers only to one point in time, which does not allow for analysing workers' skill development over time. Moreover, due to lack of data, the authors cannot explore the contribution of informal learning to skill formation.

One of the main tenets of our chapter is that the skills individuals can accumulate by learning on the job are acquired not only by formal training but also by informal learning. Early support for this idea can be found in (Mincer, 1974), who claimed that on-the-job informal learning probably constitutes the essential part and the major human capital investment in the workplace. Empirical studies such as those of Levitt, List, Syverson, et al. (2012) and Destré et al. (2008) have, respectively, shown that learning by doing and learning from others are significantly important in explaining workers' earnings and firm productivity. However, due to data limitations and the assumption in standard models that experience absorbs the work-related learning effect, there is hardly any empirical evidence that informal learning on the job is positively related to workers' skill improvement or productivity.

Last but not least, human capital literature indicates that there could be direct complementarity and cross-productivity (Cunha and Heckman, 2007; Heckman and Carneiro, 2003) between training and informal learning investments in the workplace. If training participation encourages informal learning and vice versa, investments in one type of learning could raise the marginal productivity of investments in the other type, in terms of skill accumulation. That is, skills acquired by training and informal learning could boost each other and thus further reinforce the skill development of workers who participate in both types of work-related learning. In our study, we explore the probability of such complementarity and cross-productivity of work-related training playing a part in the skill development of European employees. Hence, the availability of measures of training participation and informal learning as well as skill changes enables us, to some extent, to open the black box in the economic literature and provide additional insight into the transfer of lifelong learning to workers' skills (De Corte, 2003; De Grip and Sauermann, 2013).

5.2.2 Skill mismatch and Human Capital Investment

Research on job mismatch concentrates mostly on the wage outcomes of overeducation (see Chevalier, 2003; Di Pietro and Urwin, 2006; Dolton and Silles, 2008; Groot, 1996; Hartog, 2000; McGuinness, 2006). More recently, the literature has exhibited a shift in emphasis from overeducation to skill mismatch (see Chevalier and Lindley, 2009; Green and Zhu, 2010; Mavromaras and McGuinness, 2012; Mavromaras, McGuinness, O'leary, Sloane, and Fok, 2010; McGuinness and Byrne, 2014; McGuinness and Sloane, 2011; McGuinness and Wooden, 2009; O'Leary, Sloane, McGuinness, O'Connor, and Mavromaras, 2009). These studies have shown that overeducation and overskilling refer to different phenomena and that overeducation may not fully capture the extent to which an individual's skills are utilised at work. Educational attainment does not incorporate any measure of ability³ or skills acquired through employment, while job entry requirements are imprecise at measuring a job's skill content. Measuring

³ It has been argued that overeducated workers are likely of lower ability and, therefore, the wage penalty could be largely explained by this unobserved heterogeneity (Groot, 1996; Sloane, Battu, and Seaman, 1999). This supports the idea that employers learn about the productive abilities of overeducated employees and pay them lower wages.

workers' skill mismatch could solve these difficulties by requiring individuals to compare the actual skills requirement of their job with their own skills acquired by either initial education, or training, or informal learning or related to their general ability.

On the one hand, the objective measure of educational mismatch seems statistically superior. However, apart from being rarely available at the necessary job level, these indicators rely on the symmetry in the distribution of required education along occupations, and have been shown to be rather poorly correlated with skill mismatches (Allen and Van der Velden, 2001; Green and McIntosh, 2007; Sloane, Battu, and Seaman, 1996). On the other hand, self-reported measures of skill mismatch avoid the symmetry assumption and incorporate more than only formal education but rely on subjective assessments. Although susceptible to measurement error due to subjective bias, skill mismatch is still considered a more adequate and potentially more robust measure of skills underutilisation and overutilisation than educational mismatch is (Linsley et al., 2005; Mavromaras and McGuinness, 2012).

The literature on educational and skill mismatches coexists with the empirical literature on wages and training; however, there has been little interaction between the two fields, despite the fact that on-the-job human capital investments could be a response to initial conditions of job–skill mismatch. Economic theory implies this connection in search and matching models in which training is considered a supplement to formal education in the way that it bridges the gap between generic skills acquired through schooling and specific skills required in the workplace (Acemoglu and Pischke, 1999b; Arulampalam et al., 2004). Consequently, training is expected to contribute to the adjustment between workers' potential productivity and the productivity ceiling of the job in which they are employed (Blázquez and Jansen, 2008).

Models of skill formation have also underlined that, besides skill augmentation, training also plays a critical role in the restoration and replenishment of human capital in the context of skill depreciation (Mincer and Ofek, 1982; Rosen, 1975). In this regard, empirical studies find that overeducated workers participate less often in training than those who are well matched, whereas undereducated workers participate more often (Messinis and Olekalns, 2007; Van Smoorenburg and Van der Velden, 2000). These studies have also found that training helps to close the gap between the actual and required education of undereducated workers through the acquisition of new skills and that training offsets the depreciation of the human capital of overeducated workers. Messinis and Olekalns (2008) find, moreover, that training participation relates to substantial wage benefits for undereducated workers in relation to their well-matched co-workers and enables overeducated employees to reduce the wage penalty associated with the mismatch. The authors also find that these wage returns to training are asymmetric, that is, the undereducated seem to benefit more than those with just the right education, while the overeducated benefit the least.

The question whether the relationship between work-related learning and workers' skill improvement differs by their initial mismatch status has not been analysed in the empirical literature. A mismatch between employees' skills and

the level of their jobs could have important implications on skill returns to investment in both training and informal learning due to several reasons. In comparison with workers in a well-matching job, underskilled and overskilled workers likely have different aims when participating in job-related learning. The former probably need more training and learning to perform at an adequate level in such challenging jobs, whereas the latter may be more likely to invest in learning to keep their skills up-to-date or to counteract skill depreciation, which might not reveal skill improvement *per se*.

In this respect, our study relates to the literature on human capital depreciation, which has also explored the between skill mismatch and skill obsolescence. Building on the ‘use it or lose it’ hypothesis or the ‘atrophy’ of a worker’s skills by non-use, this literature has shown that overqualified workers who are unable to fully utilise their skills in their job are less able to sustain their cognitive level and are, therefore, more vulnerable to experience skill decline (e.g. De Grip, Bosma, Willems, and Van Boxtel, 2008; De Grip and Van Loo, 2002).

If we integrate the preceding evidence on skill depreciation into the standard model of skill formation,⁴ the skill returns of further investments in training and informal learning are expected to vary with the workers’ job–skill mismatch status. A key feature of the technology of skill development is that skills bolster each other over the life cycle because acquired skills raise both the level and productivity of subsequent learning investments in a multiplier process (Cunha and Heckman, 2007). However, these attributes of human capital might be reduced if skills are underutilised in the job or increased if job skill demands are higher. Being underskilled at job entry could serve as a positive context, not only in raising the level of further learning investments but also in the effectiveness of these investments in human capital, given their initial stock of skills. Conversely, being overskilled at job entry could act as a negative context, not only because a worker’s initial stock of skills may deteriorate due to non-use, but also because this deterioration may decrease the effectiveness and complementarity of further learning investments in the accumulation of human capital. This is consistent with Cunha and Heckman (2007) evidence that higher initial stocks of human capital are not productive if they are not followed up by later investments over the life cycle.

This notion of the skill multiplier presented by Cunha and Heckman (2007) enable us to better understand the interaction between work-related human capital investments and the degree of on-the-job utilisation of the initial stock of workers’ skills. In comparison with individuals whose acquired skills are appropriately used in their job, those with a higher initial stock of skills than required for their job are expected to have greater skill multiplier potential but, at the same time, lower levels of investment in training and informal learning on the job. According to the technology of skill formation, this implies a lower complementarity of human capital investments and, therefore, a lower marginal productivity of later investments. Conversely, underskilled workers are expected to have a lower skill multiplier potential because of their lower stock of skills compared to what is required in their jobs. These workers could, however, show

⁴ Models on skill formation such as those in (Cunha and Heckman, 2007; Killingsworth, 1982) assume that skills acquired in one period persist into the future without any deterioration.

greater investments in work-related training and informal learning to substitute for their lack of skills needed to perform at an adequate level in their job, which could produce stronger complementarity between human capital investments and hence a higher marginal productivity of later remedial investments in training and informal learning.

There is one more reason to expect that matched and mismatched workers differ in the extent to which their skill development and learning investments are related. The notion that worker motivation fosters a more efficient cognitive acquisition of skills suggests that, among overskilled workers, the productivity of training and learning on the job may be lower if they are generally less motivated and satisfied, as has been shown in the literature (e.g. Green and Zhu, 2010; Hersch, 1991; Verhaest and Omey, 2009). We build on the theoretical and empirical studies discussed above to explore in more detail the heterogeneity in the relationship between work-related learning and skill development with regard to employees' skill mismatch at job entry.

5.3 Data and Descriptive Analyses

5.3.1 Data and Sample

We use data from the European Skills and Jobs Survey, conducted in 2014 by Cedefop in 28 European countries.⁵ The survey was based on a representative sample of the working population aged 24–65 years in each of the participant countries and administered either online or by telephone to 48,676 individuals. The survey provides unique self-reported measures of employees' changes in skill accumulation as well as changes in skill mismatch over years of tenure with the same employer. Comparable measures are not available from any other large-scale dataset. Furthermore, this survey provides information on both the incidence of training and the intensity of informal learning in the workplace, in addition to other individual, job, and employer characteristics.⁶ We restrict our analyses to full-time employees⁷, obtaining a sample of 37,285 individuals. Table 5.A1 in the Appendix shows the sample distribution by country.

5.3.2 Variables and Descriptive Analyses

Table 5.A2 in the Appendix shows the main descriptive statistics of the variables included in our analyses.

⁵European Centre for the Development of Vocational Training (Cedefop). Cedefop European skills and jobs survey (ESJS), Wave 1, Spring 2014 [computer file], 1st edition, Thessaloniki: Greece, Downloaded from: <http://www.cedefop.europa.eu/en/events-and-projects/projects/analysing-skill-mismatch>. ESJS microdata are Cedefop copyright and are reproduced with the permission of Cedefop. Further information is available at Cedefop (2015).

⁶ This dataset has, however, some limitations for our analyses, which we discuss in Section 5.5.1.2 when performing robustness checks of our main results.

⁷ We consider those who reported a minimum of 35 working hours a week full-time employees.

5.3.2.1 Dependent Variable

Our outcome variable, workers' skill development is based on self-assessed changes in skills⁸ since the start of their current job. It is derived from the following question:⁹

Compared to when you started your job with your current employer, would you say your skills have now improved, worsened, or stayed the same? Please use a scale of 0 to 10, where 0 means your skills have worsened a lot, 5 means they have stayed the same, and 10 meansdht tadjstwhy they have improved a lot.

The mean reported skill development is 7.76, with a standard deviation of 1.77. As shown in Table 5.1, approximately 86 percent of the individuals in the sample reported that their skills had improved (scores of six to 10), whereas only 14 percent indicated that their skills had stayed the same (score of five) or worsened (scores of one to four).

Table 5.1: Distribution of skill development since job entry

Skill change	%	
My skills have worsened a lot	0	0.2
	1	0.2
	2	0.5
	3	0.8
	4	1.3
My skills have stayed the same	5	10.9
	6	7.5
	7	16.9
	8	25
	9	17.1
My skills have improved a lot	10	19.7

5.3.2.2 Explanatory variables

First, we distinguish between two types of work-related learning: training and informal learning.

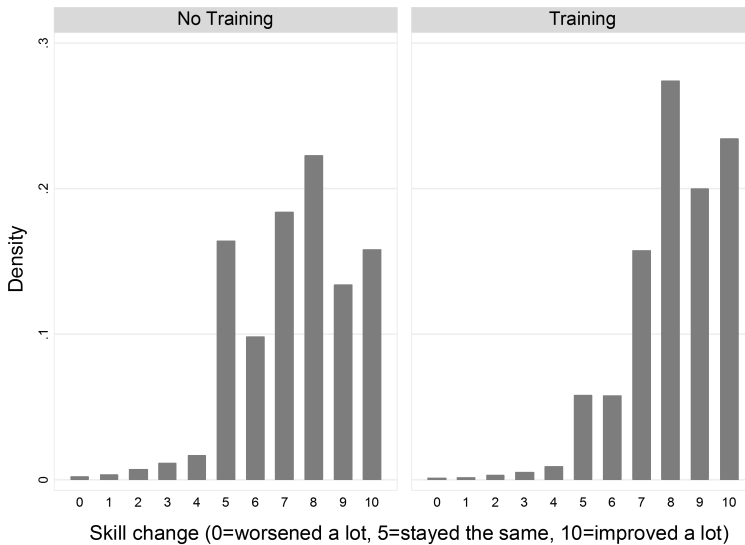
The variable Training is a dummy variable that takes the value one if the employee has participated in training courses since the start of the current job and zero otherwise. It is based on the question, 'Since you started your job with your current employer, have you attended training courses (work-based, classroom-based, and online)?' Since this question was only asked to those who reported having experienced positive skill development, we use the response to the question on training participation in the last 12 months for those whose skill

⁸ Skills were defined in the questionnaire as 'all of the knowledge, abilities, and competences that you have gained as part of your education and also during the time you have been working'.

⁹ The response rate to this question was 98 percent; only 2 percent of employees stated they had current skills not comparable to those they had before or did not to know the answer to the question.

declined.¹⁰ Table 5.A2 of the Appendix shows that 62 percent of all employees in our sample participated in training courses at least once since they started their current job, while 57 percent did so during the last 12 months. Among the latter, we observe that 45 percent underwent their training mainly during working hours, while 22 percent did so outside working hours. For 69 percent of the workers, training was fully financed by the employer, whereas 12 percent financed it themselves. As shown in Figure 5.1, the density distribution of employees' development of skills shifts to the right when workers participate in training. We also observe this significant difference in the mean value of the variable for skill development by training participation, which is 8.40 for training participants and 7.07 for non-participants. This already indicates a positive relation between training participation and skill development.

Figure 5.1: Skill development distribution by training participation.

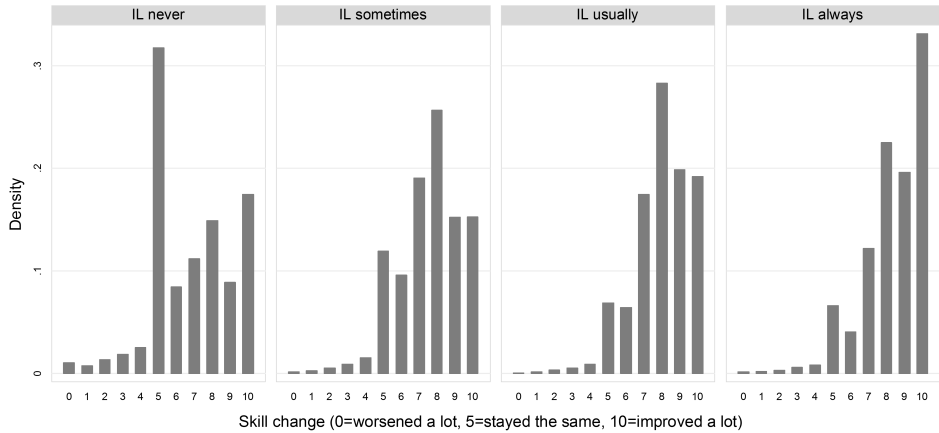


Informal learning is measured by a categorical variable IL derived from the question ‘How often, if at all, does your job involve learning new things?’ The respondent’s options were ‘never’, ‘sometimes’, ‘usually’, or ‘always’. Table 5.A2 of the Appendix shows that 55 percent of all the employees in our sample stated that they usually or always learned informally on the job, whereas only 4 percent said they were never involved in learning at work. Importantly, as shown in Figure 5.2, the density distribution of skill improvement is more concentrated to the right when workers are more often involved in informal learning. Accordingly, the mean value of the variable skill development also varies with the frequency of informal learning on the job. These values are 8.35 and 6.69

¹⁰ Training participation since the start of the job and during the last 12 months are highly positively correlated (0.67). A total of 81 percent of workers who answered both questions on training participation since job entry and during the last 12 months gave the same answer to both questions.

for workers who stated that they were always and never involved in informal learning while working, respectively. This provides first evidence that informal learning is also positively related to workers' skill development. In additional analyses, we differentiate between three types of informal learning by including dummy variables for 1) learning from colleagues and supervisors, 2) learning by trial and error, and 3) learning from self-study.¹¹

Figure 5.2: Skill development distribution by frequency of informal learning.



Second, we distinguish between workers who experienced a mismatch at the start of their current job and those who did not. Initial job–skill mismatch status is a categorical variable that takes three different values (*initially well matched*, *initially underskilled*, *initially overskilled*) corresponding to the three possible responses to the following question:

When you started your job with your current employer, overall, how would you best describe your skills in relation to what was required to do your job at that time? a) my skills were matched to what was required by my job, b) some of my skills were lower than what was required by my job and needed to be further developed), or c) my skills were higher than required by my job.

In our sample, 51 percent of all the employees stated they had a good skills match at the start of their jobs, while 24 percent considered themselves initially underskilled and 25 percent considered themselves initially overskilled. As shown in Figure 5.3, the distribution of skill development differs between the three

¹¹ These variables are based on the question ‘Since you started your job with your current employer, have you done any of the following...?’ Respondents could indicate as many of the following answers as applicable: ‘a) your supervisor taught you on-the-job, b) you learned by interacting with colleagues at work, c) you learned at work through trial and error, and d) you learned by yourself (e.g. with the aid of manuals, books, videos or on-line materials)’.

different groups in favour of employees who were initially underskilled. We also observe significant differences in the mean value of the variable for skill development by skill mismatch status, which is 7.78 for the well matched, 8.41 for the underskilled, and 7.15 for the overskilled. This descriptive evidence suggests that workers who start a job with fewer skills than required make the largest skills progress when gaining years of tenure.

Figure 5.3: Skill development distribution by initial job–skill mismatch.

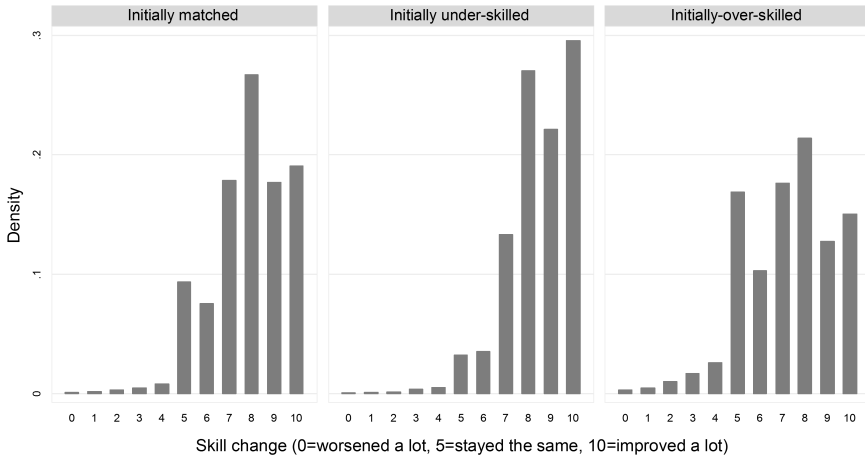


Table 5.A2 shows other differences between initially underskilled and overskilled workers. For those who were initially overskilled, a slightly higher proportion is male and higher educated. Moreover, these workers more often have temporary contracts and fewer years of tenure. Furthermore, the table shows a higher percentage of underskilled workers in manufacturing and among professionals, technicians, and workers in crafts and related trades. Overskilled workers are overrepresented in the sales and transportation industries, as well as in the service and sales, and clerical support occupations. It is worth mentioning that there is no difference in workers’ ages between the three skill mismatch groups (mean= 42, standard deviation = 9.8) or in the sizes of the firms that employ them.

It is important to note that underskilled workers participated more often in training and formal education. They also stated they learned more often on an informal basis than well-matched and overskilled workers did. The latter invested the least in their human capital.¹² This may not be surprising, since

¹² Multivariate analyses support these correlations. We find that initially underskilled workers are, on average, 6.8 percentage points more likely to participate in training courses, 4.8 percentage points more likely to engage in informal learning, and more intensively involved in informal learning than well-matched workers by approximately 0.55 of a standard deviation. Conversely, the initially overskilled have a lower probability of training participation (-1.4 percentage points) and a lower probability of taking part in informal learning (-1.6 percentage points) in comparison with workers who started in a well-matched job. In addition, informal learning intensity is also lower among the overskilled than among well-matched workers by approximately 0.94 of a standard deviation (see Table 5.A4 in the Appendix).

these workers already had more skills than required in their job. This suggests that having a job that initially mismatches the skills of workers is related to participation in training as well as informal learning, which could influence workers' skill development.

5.3.2.3 Control variables

We explicitly include participation in formal education – which led to a higher degree while working for the current employer – as a control variable in our model. This variable measures human capital investments in the form of schooling rather than job-related learning. Formal education is a dummy variable for participation in formal educational programmes resulting in a higher or different educational degree while working for the current employer. We construct this variable by assigning the value one for those who achieved their highest level of education after they started to work with their current employer and zero otherwise. As shown in Table 5.A2, 14 percent of all employees in our sample participated in formal education while working for their current employer. Formal education also seems to be positively correlated to skill development, a reason to include it as a control variable.

As suggested by human capital theory, we additionally control for age, gender, educational level (low, middle, and high), tenure, type of contract (permanent, fixed-term temporary, agency temporary, and no formal contract), occupation (nine one-digit ISCO categories), industry (10 one-digit ISIC categories), firm size (five categories), and country dummies. Moreover, we include dummies to control for major changes since the start of the job in regards to job position or job task content,¹³ working methods,¹⁴ and the need to learn new things,¹⁵ which could affect the relationship of main interest in this chapter. We also include a dummy variable that indicates whether the survey was conducted by telephone.

¹³ Major changes in the job position over tenure are considered, that is, if the worker has experienced a promotion, a demotion, a change of unit/department, or a substantial change in the nature of job tasks. These dummy variables are derived from the answers to the question, 'Since you started working for your current employer, have any of the following changes in your role taken place?' The respondent's options were: 'a) I have been promoted to a higher level position, b) I moved to a different unit/department, c) I have not been promoted or moved department but the nature of my tasks and responsibilities has changed, d) I now have a lower-level position than when I started, and e) no changes, my role has remained the same'.

¹⁴ These dummy variables are derived from the answers to the question 'Since you started your main job have any of these changes taken place in your workplace?' The respondent's options taken into account were 'a) Changes to the technologies you use (e.g. machinery, ICT systems) and b) changes to your working methods and practices'.

¹⁵ This categorical variable is derived from the question 'Has the need to learn new things increased, decreased or remained the same since you started your job with your current employer? Please use a scale of 0 to 10, where 0 means it has decreased a lot, 5 means it has stayed the same and 10 means it has increased a lot'. Answers from zero to four are considered a decrease, whereas answers from six to 10 are considered an increase in the need of learning. Answer five is the base category.

5.4 Estimation Method

To estimate the relation between employees' job-related learning and skill development, we use ordered probit models. The fact that responses to our dependent variable are concentrated in some categories suggests that the meaning of certain categories is more expansive than others. In this case, OLS estimation is likely to give misleading results (Long and Freese, 2006; Winship and Mare, 1984). Therefore, we consider the self-reported measure of individuals' skill changes as an ordinal structure in which the distances between the categories are unknown and allowed to be unequal. Let SD_i denote an observable ordinal variable coded from zero to 10 based on responses to the individual skill change question described in the previous section. These choices are modelled based on an unobservable latent continuous variable (SD_i^*) that can be expressed as a function of a set of observable factors (Z_i) and unobservable factors (u_i) using the following linear relationship:

$$SD_i^* = Z_i' \beta + u_i = \gamma' X_i + \delta L_i + \zeta ISM_i + \psi(L_i * ISM_i) + u_i \quad (5.1)$$

where X is a vector of covariates composed of worker and firm characteristics along with a set of country dummies, L is a vector of participation in training and informal learning variables, ISM is an indicator of the initial job-skill match, and $u_i \approx N(0, 1)$. The existence of a set of $K - 1$ ordered threshold parameters is also assumed, such that the individual responds with category k if and only if $SD_i^* \in [\theta_{k-1}, \theta_k]$. In general terms, we can write $Prob(SD_i = k | Z_i) = \Phi(\theta_k - Z_i' \beta) - \Phi(\theta_{k-1} - Z_i' \beta)$ for $k = 0, \dots, K$, where $\Phi(\cdot)$ denotes the cumulative distribution function of u_i for the standard normal. The first and final intervals are open ended, so for $k = 0, \Phi(\theta_{k-1}) = \Phi(-\infty) = 0$ and for $k = 10, \Phi(\theta_k) = \Phi(+\infty) = 1$. The regression parameters γ, δ, ζ , and ψ and the $K - 1$ threshold parameters are obtained by maximising the log-likelihood function subject to $\theta_k > \theta_{k-1}$ for all k . We use a robust clustered estimator of variance to allow for intragroup correlation at the country level (Wooldridge, 2010).

As described above, our analyses consider interactions between the learning variables L (training, informal learning, and formal education) and the employee's initial skills match ISM . As Greene (2010); Karaca-Mandic, Norton, and Dowd (2012); Norton, Wang, Ai, et al. (2004) have shown, the interpretation of interaction terms in linear models does not extend to nonlinear models. Basically, the interaction effect in nonlinear models cannot be evaluated by looking at the sign, magnitude, or statistical significance of the coefficient of the interaction term (Ai and Norton, 2003). For nonlinear models that include interactions between categorical variables as in this chapter, the interaction effect becomes the following discrete double difference:

$$\frac{\Delta^2 \Phi(Z' \beta)}{\Delta L * \Delta ISM} = \frac{\Delta \{ \Phi[\delta + \zeta ISM + \psi(L * ISM) + \gamma' X] - \Phi[\zeta ISM + \gamma' X] \}}{\Delta ISM} \quad (5.2)$$

$$\frac{\Delta^2 \Phi(Z' \beta)}{\Delta L * \Delta ISM} = \Phi(\delta + \zeta + \psi + \gamma' X) - \Phi(\delta + \gamma' X) - \Phi(\zeta + \gamma' X) + \Phi(\gamma' X)^2$$

Some implications need to be taken into account. First, the interaction effects in nonlinear models are conditional on the independent variables. Second, since the additive terms can be either positive or negative, the interaction effects could have opposite signs for different values of covariates and, therefore, the sign of ψ does not necessarily reflect the sign of the interaction effects. Third, even if ψ is zero, the interaction effects could be nonzero. Finally, the statistical significance tests of the interaction terms need to be associated with the entire double difference (Ai and Norton, 2003; Norton et al., 2004). Taking these implications into account, we compute and report, as suggested by Karaca-Mandic et al. (2012); Long and Freese (2006), full interaction marginal effects (cross-differences) and their statistical significance to correctly interpret our results.

5.5 Estimation Results

5.5.1 Work-Related Learning and Skill Development

We estimate an ordered probit regression for skill development that includes interactions between the work-related learning variables and the employee's initial skills match, as described in the previous section. The resulting coefficients, however, are not directly interpretable (Greene, 2010; Long, 1997; Long and Freese, 2006). We therefore provide in Table 5.2 the corresponding average marginal effects (AMEs).¹⁶ To facilitate the interpretation of results, we compute AMEs in four categories: worsened skills (scores zero to four), no or hardly any change in skills (scores five and six), intermediate improvement of skills (scores seven and eight), and high improvement of skills (scores nine and 10).¹⁷

In line with expectations from human capital theory, the results in Table 5.2 show that both participation in training and informal learning are significantly and positively associated with employees' skill development. These results also show that the AME of most of the explanatory variables related to the probability of high skill improvement is crucial in the way that it is offset by the distinctive probabilities of being in the other categories, zero to eight. These marginal effects confirm our descriptive results that the probability of high skill improvement is greater for employees who participated in training or informal learning in comparison with those who did not. More precisely, workers who participated in training are, on average, 8.3 percentage points more likely to have highly improved their skills than those who did not participate in any training course. Similarly, participation in training seems to reduce the odds of experiencing skill worsening and stagnation by 1.4 and 4.9 percentage points, respectively. In addition, employees' involvement in informal learning is also

¹⁶ The coefficients are shown in Table 5.A3

¹⁷ According to Long and Freese (2014), having more than two outcomes creates a challenge in summarising the effects of the independent variables in a way that fully reflects key substantive processes without overwhelming and distracting detail. We compute marginal effects in the four categories mentioned based on the criteria that the probabilities in the same group are of the same sign and similar size.

Table 5.2: Average marginal effects (AMEs) on workers' skill development

Skill change	0-4	5-6	7-8	9-10
Training	-0.014*** (0.001)	-0.049*** (0.004)	-0.020*** (0.003)	0.083*** (0.007)
IL sometimes	-0.012** (0.004)	-0.025** (0.011)	0.003 (0.004)	0.034** (0.016)
IL usually	-0.021*** (0.005)	-0.056*** (0.012)	-0.009** (0.004)	0.086*** (0.012)
IL always	-0.029*** (0.005)	-0.099*** (0.012)	-0.040*** (0.006)	0.168*** (0.016)
Formal education	-0.008*** (0.001)	-0.030*** (0.004)	-0.015*** (0.003)	0.053*** (0.007)
Initially underskilled	-0.012*** (0.001)	-0.056*** (0.003)	-0.041*** (0.003)	0.109*** (0.006)
Initially overskilled	0.015*** (0.002)	0.044*** (0.006)	0.007** (0.002)	-0.066*** (0.008)
Age	0.000*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	-0.002*** (0.000)
Female	-0.010*** (0.001)	-0.039*** (0.004)	-0.021*** (0.002)	0.070*** (0.006)
Intermediate level education	0.003*** (0.001)	0.013*** (0.004)	0.009** (0.003)	-0.025*** (0.009)
High level education	0.010*** (0.002)	0.038*** (0.005)	0.023*** (0.003)	-0.071*** (0.010)
Years of tenure	-0.001*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	0.006*** (0.000)

Notes: This table shows AMEs computed on the ordered probit specification (2) in Table 5.A3 in the Appendix. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 37, 285$.

positively related to their skills improving. For instance, the likelihood of a high improvement of skills is 16.8, 8.6, and 3.4 percentage points greater for workers who, respectively, always, usually, and sometimes learn informally on the job, in comparison with those who are never involved in informal learning in their job.

Generally, our results so far suggest that the contribution of informal learning to workers' skill development seems to be larger than that of training participation. To provide more insight into this possibility, we perform an estimation that includes a dummy variable for informal learning¹⁸ instead of the original categorical variable. The predicted probabilities show that, indeed, the relationship of skill development with informal learning on the job is stronger and significantly different than that with training participation (see Panel A of Table 5.A5 in the Appendix). This supports Mincer's (1974) claim that informal learning probably constitutes the major part of human capital investment in the

¹⁸ This variable takes the value zero if the employee claimed to never be involved in informal learning on the job and takes the value one if the employee claimed to be involved in informal learning either sometimes, usually, or always.

workplace, given that workers likely spend much more time learning by doing in their job than in other training activities.

The results in Table 5.2 also suggest that the initial job–skill mismatch of workers significantly helps to explain their skill development while in the same job. We find that initially underskilled workers develop their skills more than those who started in a job that matched their skills well. Conversely, over-skilled employees are more likely to experience skill worsening (by 1.5 percentage points) and stagnation (by 4.4 percentage points) than well-matched employees, confirming the evidence on skill depreciation shown by De Grip et al. (2008).

Regarding the covariates in our model, we find that the marginal probability of workers’ skill development decreases with age and is lower for employees who are more educated, for those who have temporary or agency contracts instead of permanent contracts, and for individuals who have experienced a decrease in the need to learn new things since the start of their job. Conversely, it increases with participation in formal education and years of tenure (which compensates for the negative effect of age by approximately three times) and tends to be higher for female employees and for those who have experienced an increasing need for learning or a major change in their job position, job task content, or working process. Other controls indicate that high skill development is less likely for individuals employed in low-skilled occupations.

5.5.1.1 Heterogeneity by Skill Mismatch Status at Job Entry

As explained in Section 5.4, we include interaction terms in our main ordered probit model to see whether there is heterogeneity in the relation between job-related learning and workers’ skill improvement due to differences in their initial job–skill mismatch status. Since the interpretation of these interactions is not as straightforward as it is for linear models, we compute the corresponding discrete double differences and statistical significance within and between the different skill mismatch statuses of workers. Two types of heterogeneity can be analysed. First, in Table 5.3, we show the difference in skill development of those who have been engaged in training or informal learning and those who have not within the same initial job–skill mismatch group.

Table 5.3 shows that the general findings of Table 5.2, that those who participate in training or informal learning show a larger skill development, hold for all workers, independent of their initial skill mismatch status. Compared to workers with the same initial skill mismatch status, those who participated in training or informal learning are more likely to improve their skills than those who have not been involved in any learning activity. Nonetheless, there seems to be some heterogeneity in the magnitude of these relationships, particularly among workers with a job above their skill level. We find that these latter individuals do not show any significant difference in their skill improvement if they never participate in informal learning or if they do so only occasionally. Moreover, the association between their skill improvement and informal learning on a usual basis is shown to be weaker than – and significant with only 90 percent of confidence – within the well-matched and overskilled groups. This result suggests that informal learning needs to take place more often among underskilled

employees to increase their probability of skill progress. Further analyses show that, within this group of underskilled employees, the relation of skill development with training is actually greater than that with informal learning, which is exactly the opposite within workers in a well-matching job or a job below their skill level (see Panel B of Table 5.A5 in the Appendix).

Table 5.3: AMEs of work-related learning on workers' skill development within job-skill (mis)match groups

Skill change	0-4	5-6	7-8	9-10
Initially well matched				
Training	-0.012*** (0.001)	-0.051*** (0.004)	-0.022*** (0.002)	0.085*** (0.007)
IL sometimes	-0.008** (0.003)	-0.026** (0.011)	-0.005*** (0.002)	0.039** (0.020)
IL usually	-0.015*** (0.005)	-0.055*** (0.015)	-0.017*** (0.003)	0.087*** (0.015)
IL always	-0.023*** (0.005)	-0.099*** (0.015)	-0.050*** (0.005)	0.172*** (0.018)
Initially underskilled				
Training	-0.006*** (0.001)	-0.035*** (0.007)	-0.038*** (0.006)	0.079*** (0.014)
IL sometimes	-0.001 (0.003)	-0.002 (0.014)	-0.001 (0.013)	0.004 (0.030)
IL usually	-0.004* (0.002)	-0.023* (0.012)	-0.022* (0.012)	0.049* (0.028)
IL always	-0.008*** (0.003)	-0.055*** (0.016)	-0.070*** (0.016)	0.133*** (0.021)
Initially overskilled				
Training	-0.021*** (0.002)	-0.063*** (0.006)	-0.005** (0.002)	0.089*** (0.008)
IL sometimes	-0.023** (0.009)	-0.051*** (0.016)	0.013* (0.007)	0.061*** (0.012)
IL usually	-0.038*** (0.010)	-0.096*** (0.017)	0.010* (0.005)	0.124*** (0.014)
IL always	-0.051*** (0.010)	-0.145*** (0.018)	-0.010* (0.005)	0.206*** (0.017)

Notes: This table shows AMEs computed on the ordered probit specification (2) in Table 5.A3 in the Appendix. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 37, 285$.

Remarkably, Table 5.3 indicates that also among the initially overskilled employees, training and informal learning are positively associated with their skill development. For instance, overskilled workers who participate in training or always engage in informal learning are, respectively, 8.9 and 20.6 percentage points more likely to develop their skills highly than overskilled workers who do not participate in training or who never engage in informal learning on the job. This could be because overskilled employees who invest in the development of their human capital acquire new skills that are different from those they have

previously accumulated (e.g. non-technical or non-cognitive skills) or are more functional in offsetting skill depreciation. The latter explanation could be inferred from the significantly larger probabilities for overskilled workers in the skill change categories scored zero to four, and five and six (i.e. skill decline and more or less stable skills) in all types of learning.

Second, in Table 5.4, we show the actual interaction effects, that is, the differences in skill development between well-matched and underskilled employees who participated in work-related learning, on the one hand, and between well-matched and overskilled employees who participated in work-related learning, on the other hand. These are the actual interaction results, which need to be interpreted simultaneously with the AMEs for well-matched workers, the reference category, shown in the first panel of Table 5.3.

Table 5.4: AMEs of work-related learning on workers' skill development between job-skill s (mis)match groups (Interaction effects)

Skill change	0-4	5-6	7-8	9-10
Initially underskilled (well matched ref.)				
Training	-0.009*** (0.001)	-0.050*** (0.003)	-0.048*** (0.004)	0.107*** (0.007)
IL sometimes	-0.014*** (0.001)	-0.060*** (0.004)	-0.031*** (0.002)	0.105*** (0.006)
IL usually	-0.006*** (0.001)	-0.044*** (0.004)	-0.059*** (0.007)	0.109*** (0.011)
IL always	-0.011*** (0.001)	-0.056*** (0.005)	-0.045*** (0.004)	0.112*** (0.009)
Initially overskilled (well matched ref.)				
Training	0.011*** (0.001)	0.040*** (0.005)	0.015*** (0.002)	-0.066*** (0.008)
IL sometimes	0.019*** (0.003)	0.054*** (0.008)	0.002 (0.001)	-0.075*** (0.009)
IL usually	0.007*** (0.002)	0.033*** (0.007)	0.022*** (0.004)	-0.062*** (0.008)
IL always	0.010*** (0.001)	0.037*** (0.005)	0.011*** (0.002)	-0.058*** (0.006)

Notes: This table shows AMEs computed on the ordered probit specification (2) in Table 5.A3 in the Appendix. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 37, 285$.

Altogether, these results clearly show differences between workers who started in a job that matched their skills well and those who started in a job at a level either above or below their own skill level. In comparison with well-matched employees, the strength of the relationship between workers' skill development and their participation in training or informal learning is stronger for those who were initially underskilled and weaker for those who were initially overskilled. For instance, underskilled employees who participated in training are 10.7 percentage points more likely to be in the two highest categories of skill development (nine

to 10) than well-matched workers with similar learning investments. Similarly, underskilled workers who are always learning informally on the job are 11.2 percentage points more likely to improve their skills highly. This suggests that, for underskilled workers, who are probably employed in a more demanding job, participation in training or informal learning on the job is more favourable for their skill development. This could be due to richer learning opportunities at work and greater interest in maintaining their jobs (De Grip et al., 2008).

Conversely, overskilled employees who participated in training are, on average, 6.6 percentage points less likely to be in the two highest categories of skill development than similar workers in a well-matching job are. This implies, at the same time, that overskilled employees who participate in training are 1.1 percentage points more likely to experience skill depreciation (category zero to four) and 4 percentage points more likely to face stagnation in their skills (category five to six) than trained matched workers are. We find a similar pattern for informal learning. Compared to well-matched workers with similar learning investments, overskilled employees who report that they always learn informally in their job are 5.8 percentage points less likely to improve their skills highly, which implies that, at the same time, they are 0.1 percentage points more likely to experience skill worsening and 3.7 percentage points more likely to face skill stagnation.

This does not mean that learning has a negative impact on the skill accumulation of overskilled workers but that the positive relation is much weaker than it is for workers in a well-matching job. For example, whereas well-matched employees who participate in training and who are always learning informally have average probabilities of high skill development of 8.5 percent and 17.2 percent, respectively, the same probabilities for overskilled employees are, on average, 1.9 (8.5 minus 6.6) percent and 11.4 (17.2 minus 5.8) percent, respectively. This result suggests that training and informal learning on the job are much less beneficial for overskilled workers than for those who are employed in a well-matching job. Nonetheless, the more often overskilled workers engage in training or informal learning, the lower their probability of skill decline and stagnation. This result again suggests that the learning investments of overskilled workers are more functional in preventing skill depreciation than in fostering skill accumulation.

5.5.1.2 Robustness of the Main Results

Whereas the data we use are unique in their wide range of information about workers' skill development, learning participation, and initial mismatch status, there are also some drawbacks that could affect the precision of our estimates. In this section, we discuss the issues of endogeneity (omitted variables and measurement error) and the use and timing of self-reported measures as possible sources of bias. We argue that the probability of these potential sources of bias undermining the inferences and plausibility of the empirical findings of our study is very low.

Endogeneity problems

The most important concern is the endogeneity in our model due to omitted variable bias and possible measurement error. First, unobservable characteristics such as ability and motivation may affect employees' learning participation decisions as well as their own perceived levels of skill improvement. For instance, if highly motivated workers are more likely to participate in training and informal learning and also tend to see greater improvement in their skill levels, our results would be affected by a positive bias. However, the ability bias is expected to be negative, since more able individuals will tend to invest more in their human capital on the job but will also observe generally smaller marginal increases in their skills, given the diminishing marginal utility of these investments. This would mean that our estimation results show a lower bound of the true effects.

Omitted variable bias might also arise because workers' own assessment of the (mis)match situation at job entry could vary with their personality traits, such as their degree of self-confidence. For example, individuals who are more confident about themselves will less often report an initial underskilled mismatch but might, simultaneously, be more likely to overestimate their own skill improvement. If these hypotheses are correct, then our results underestimate the true effects. Similarly, if individuals of higher ability are typically selected into well-matching or skill-challenging jobs, the expected endogeneity bias will also be negative. The opposite, however, would be expected with respect to omitted information on motivation.

Although the preceding analysis suggests more reasons to consider a negative bias for our results, it is not obvious how all these sources of endogeneity will balance out in an empirical approach. To provide some insight into these issues, we perform robustness analyses, taking into account the information on worker attitudes that is available in our dataset, that is, (i) learning motivation and (ii) the importance of career development opportunities for respondents in the acceptance of their current job.¹⁹ Table 5.5 presents the main results.

Panel A of Table 5.5 shows that, as expected, learning motivation is significantly and positively related to workers' skill development. However, including this confounder does not substantially change our main estimation results, which remain significant and of similar magnitude as those in Table 5.2.²⁰ Similarly, the estimation results in Panel B show that the importance workers gave to career development opportunities in accepting their current job is also positively associated with their skill development. Nevertheless, the predicted probabilities

¹⁹ The variable for learning motivation is a standardised index derived from the question 'On a scale from 0 to 10, where 0 is strongly disagree and 10 strongly agree, please indicate to what extent do you agree or disagree with the following statements: a) I enjoy learning for its own sake, b) I try to relate learning to practical issues, and c) I prefer to have others plan my learning'. The variable for career development opportunities is derived from the question 'Before you started working for your current employer, how important, if at all, were the opportunities for job career development in your decision to accept the job? Please use a scale of 0 to 10 where 0 means not at all important, 5 means moderately important and 10 means essential'.

²⁰ Although this holds for all workers with low or high learning motivation, we observe that workers with stronger learning motivation show a greater skills increase in relation to both training and informal learning.

Table 5.5: AMEs of work-related learning on workers' skill development including information on workers' learning and career attitudes

Skill change	0-4	5-6	7-8	9-10
<i>A. AMEs of work-related learning including learning motivation</i>				
Training	-0.015*** (0.001)	-0.045*** (0.004)	-0.018*** (0.002)	0.078*** (0.007)
IL sometimes	-0.018*** (0.005)	-0.038*** (0.010)	0.000 (0.004)	0.056*** (0.015)
IL usually	-0.027*** (0.005)	-0.068*** (0.009)	-0.009** (0.004)	0.104*** (0.013)
IL always	-0.036*** (0.006)	-0.103*** (0.010)	-0.034*** (0.005)	0.173*** (0.015)
Underskilled	-0.013*** (0.001)	-0.056*** (0.003)	-0.044*** (0.003)	0.113*** (0.006)
Overskilled	0.020*** (0.002)	0.051*** (0.005)	0.007*** (0.002)	-0.078*** (0.007)
Learning motivation (std)	-0.013*** (0.001)	-0.041*** (0.002)	-0.020*** (0.001)	0.074*** (0.003)
<i>B. AMEs of work-related learning including importance of career development</i>				
Training	-0.013*** (0.001)	-0.045*** (0.004)	-0.019*** (0.003)	0.077*** (0.007)
IL sometimes	-0.009** (0.004)	-0.021** (0.010)	0.003 (0.004)	0.027** (0.013)
IL usually	-0.017*** (0.005)	-0.048*** (0.012)	-0.007* (0.004)	0.072*** (0.018)
IL always	-0.026*** (0.005)	-0.088*** (0.012)	-0.037*** (0.006)	0.151*** (0.020)
Underskilled	-0.013*** (0.001)	-0.058*** (0.003)	-0.042*** (0.003)	0.113*** (0.006)
Overskilled	0.013*** (0.002)	0.040*** (0.006)	0.007*** (0.002)	-0.060*** (0.008)
Importance of career (std)	-0.007*** (0.001)	-0.024*** (0.002)	-0.012*** (0.001)	0.043*** (0.003)

Notes: This table shows AMEs computed on an ordered probit specification similar to column (2) in Table 5.A3 in the Appendix that includes a variable for learning attitude (Panel A) and a variable for the importance of career development opportunities to accept the job at the first place (in Panel B). The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 29, 018$ (Panel A). $N = 37, 285$ (Panel B).

for training and informal learning remain positive and significant in this specification, indicating that our main findings remain the same when we account for workers' attitude towards career development opportunities.²¹ These results all hold for overskilled, underskilled, and well-matched employees.²²

²¹ Although this holds for all workers who responded with low or high importance to their job career development opportunities, we observe that workers who considered career opportunities more important in accepting the job show a greater skill increase in relation to both training and informal learning.

²² Detailed tables of results within and across skill mismatch statuses are available upon request.

Measurement error and the timing of self-assessed variables

A second concern is the possible bias of our results due to systematic measurement errors in the self-assessed measures we use for our analyses. Our main dependent variable, skill development, and our key moderator variable in the model, initial skill mismatch status, are subjective self-assessments of workers based on a retrospective appraisal of the period since the respondent started the current job. The use of subjective data for research in topics such as this is legitimate, particularly given the difficulty of obtaining objective indicators for these variables.

However, there are two major possible problems with these data. First, the self-assessed information could involve recall error. We consider that the magnitude of this type of measurement error is likely to differ across respondents due to the variation of time elapsed since job entry. Second, it could be possible that a worker's skill development took place before the worker participated in on-the-job learning, that is, a reverse order possibility. We attempt to capture these two potential problems by estimating our main model for workers with different tenure. The idea behind these robustness analyses is that the degree to which recall bias or reverse order might affect our estimates most likely depends on the time elapsed since the start of the job.

We therefore first estimate for each additional year of tenure the corresponding probability of high skill improvement (i.e. categories nine to 10 of our dependent variable) in relation to both training and informal learning. Figure 5.4 shows the estimation results evaluated over years of tenure. The top figure represents the AMEs of training and the bottom figure the AMEs of (sometimes and always) informal learning. We observe that, for both forms of learning, the predicted marginal probabilities of high skill improvement are, as expected, slightly increasing during the first period (six years) of tenure and then continue in a constant pattern until the later years of tenure. Nonetheless, the confidence intervals show that the point estimates for the various years of tenure are not significantly different from each other. This suggests that potential sources of bias due to systematic measurement error have not influenced our main results. This result also holds for the heterogeneity of our main findings between over-skilled, underskilled, and well-matched employees, as shown in Figure 5.A1 and Figure 5.A2 in the Appendix.

Second, we assess the accuracy of our main results by comparing them with estimation results on more selective samples of workers for whom the recall bias and/or the reverse order possibility could be considered to be at a minimum. These are workers with shorter periods of job tenure or those who have finished formal education more recently and are relatively younger. We present the corresponding results in Table 5.6. This table shows that the estimation results are similar in magnitude and significance to our main results in Table 5.2. This suggests that, although a bias due to measurement error in self-assessed variables could be important for the precision of measuring the true effects, this bias is not likely to be large enough to challenge the plausibility of our findings.

Figure 5.4: AMEs of work-related learning on workers' skill development (cat. 9-10) over years of tenure.

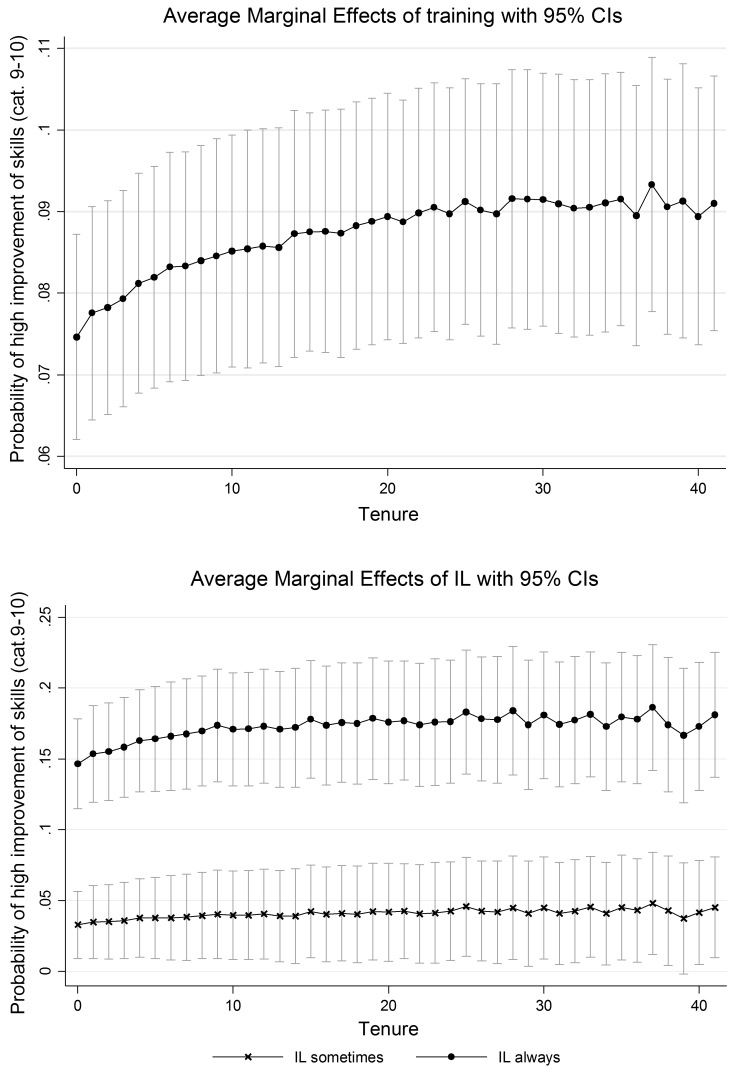


Table 5.6: AMEs of work-related learning on workers' skill development (cat. 9-10)

	(1) Tenure <= 2	(2) Tenure <= 5	(3) Tenure <= 10	(4) Left education <= 5	(5) Left education <= 10	(6) Left education <= 5 Age <= 40	(7) Left education <= 10 Age <= 40
Training	0.065*** (0.011)	0.074*** (0.008)	0.081*** (0.007)	0.080*** (0.014)	0.075*** (0.010)	0.073*** (0.015)	0.067*** (0.010)
IL sometimes	0.042** (0.019)	0.040** (0.018)	0.037** (0.017)	0.029* (0.016)	0.032** (0.015)	0.025* (0.014)	0.028** (0.013)
IL usually	0.098*** (0.017)	0.094*** (0.016)	0.089*** (0.014)	0.075*** (0.015)	0.082*** (0.016)	0.070*** (0.016)	0.077*** (0.017)
IL always	0.149*** (0.018)	0.158*** (0.018)	0.164*** (0.017)	0.154*** (0.028)	0.147*** (0.026)	0.147*** (0.028)	0.142*** (0.027)
Underskilled	0.105*** (0.012)	0.114*** (0.009)	0.119*** (0.007)	0.118*** (0.012)	0.114*** (0.009)	0.119*** (0.018)	0.117*** (0.011)
Overskilled	-0.071*** (0.009)	-0.075*** (0.008)	-0.077*** (0.008)	-0.090*** (0.017)	-0.075*** (0.010)	-0.084*** (0.017)	-0.077*** (0.012)
Tenure	0.016*** (0.004)	0.015*** (0.002)	0.012*** (0.001)	0.007*** (0.001)	0.006*** (0.001)	0.008*** (0.002)	0.007*** (0.002)
<i>N</i>	7,434	14,247	23,076	4,944	10,503	3,816	8,263

Notes: This table shows AMEs computed on an ordered probit specification similar to column (2) in Table 5.A3 in the Appendix that restricts the corresponding sample to workers with the characteristics mentioned in the column heading. Tenure refers to the time (in years) that the individual has been working for the current employer. Left education refers to the time (in years) elapsed since the worker obtained his highest educational degree. Since several people finished education while working, in columns (6) and (7) we further restrict the sample combining the criterion on time elapsed since the worker finished formal education with a criterion on age. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). This table shows the AME only for the last category, that is, high improvement of skills. The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5.1.3 Additional Evidence on Job-Related Human Capital Investments and Skill Development

In this section, we analyse whether there are any differences in the relevance of different types of training and informal learning for workers' skill development. First, we distinguish between training participation during and outside working hours, and between training paid by the employer and by the employee. Second, we distinguish between informal learning from others (co-workers and supervisors), informal learning by trial and error, and informal leaning by self-study.

Different types of training

The results in Panel A of Table 5.7 show some heterogeneity with regards to different types of training for workers with a different mismatch status. We observe that employees in a well-matching job who followed training courses during working hours show greater improvement in their skills than those who participated in training outside working hours. For underskilled employees, the contribution of training outside working hours does not seem to be significant, whereas among overskilled workers the difference between training during and outside working hours is much lower (and statistically insignificant) in comparison with the other two groups of workers. Concerning the main funding sources of training, we find that, for well-matched and underskilled employees, the contribution of training paid by the employer to their skill development seems to be larger than that of training paid by the employees themselves. Conversely, we again observe that, among overskilled workers, the difference between training paid by the employer and training paid by the employee is lower and statistically insignificant.

This heterogeneity could arise from differences in the relevance of the training content for workers with a different (mis)match.²³ Since overskilled workers participate more often than others in training outside working hours and financed by themselves, it is plausible that they consider the investment in this training to be more general and useful to improve or update skills they do not use in their job. This could provide them with more opportunities to find a better job match in their current or other firms. This could explain why, in Panel B of Table 5.7, training outside working hours and training paid by the employee seem to be as useful for the skill development of overskilled workers as it is for the well-matched, whereas training during working hours and training paid by the employer appear to be less beneficial for the former group. Lastly, Panel B of Table 5.7 also suggests that, in comparison with employees in a well-matching job, training participation seems to be more beneficial for underskilled workers, regardless of the funding source and whether it occurs during or outside working time.

²³ An alternative explanation could be that the duration of these different types of training differs across workers. We, cannot, however, analyse this possibility, since there is no information on training duration/intensity available in our dataset.

Table 5.7: AMEs of different types of training on workers' skill development

Skill change	0-4	5-6	7-8	9-10
<i>A. AMEs within job-skill (mis)match groups</i>				
Initially well matched				
Training during working hours	-0.008*** (0.001)	-0.028*** (0.003)	-0.013*** (0.002)	0.049*** (0.006)
Training outside working hours	-0.005** (0.002)	-0.008** (0.003)	-0.004* (0.002)	0.017** (0.007)
Training paid by employer	-0.006*** (0.001)	-0.031*** (0.003)	-0.012*** (0.002)	0.049*** (0.006)
Training paid by employee	-0.004* (0.002)	-0.016** (0.007)	-0.007** (0.003)	0.027** (0.012)
Initially underskilled				
Training during working hours	-0.003*** (0.001)	-0.019*** (0.004)	-0.022*** (0.004)	0.044*** (0.008)
Training outside working hours	-0.001 (0.001)	-0.005 (0.005)	-0.005 (0.005)	0.011 (0.012)
Training paid by employer	-0.003*** (0.001)	-0.020*** (0.004)	-0.022*** (0.004)	0.045*** (0.009)
Training paid by employee	-0.008** (0.004)	-0.016** (0.007)	-0.002 (0.002)	0.026** (0.013)
Initially overskilled				
Training during working hours	-0.012*** (0.002)	-0.035*** (0.005)	-0.004*** (0.001)	0.051*** (0.007)
Training outside working hours	-0.010*** (0.002)	-0.033*** (0.007)	-0.002 (0.001)	0.045*** (0.010)
Training paid by employer	-0.013*** (0.001)	-0.020** (0.009)	-0.022** (0.010)	0.055*** (0.006)
Training paid by employee	-0.011** (0.001)	-0.031*** (0.004)	-0.004*** (0.001)	0.046** (0.021)
<i>B. AMEs between job-skill (mis)match groups (well matched ref)</i>				
Initially underskilled				
Training during working hours	-0.009*** (0.001)	-0.051*** (0.003)	-0.044*** (0.003)	0.104*** (0.005)
Training outside working hours	-0.006 (0.005)	-0.022** (0.010)	-0.018*** (0.003)	0.046** (0.022)
Training paid by employer	-0.008*** (0.001)	-0.053*** (0.004)	-0.043*** (0.004)	0.104*** (0.008)
Training paid by employee	-0.010*** (0.001)	-0.052*** (0.006)	-0.037*** (0.007)	0.099*** (0.014)
Initially overskilled				
Training during working hours	0.009*** (0.001)	0.019*** (0.004)	0.012*** (0.002)	-0.040*** (0.008)
Training outside working hours	0.01 (0.010)	0.016 (0.010)	0.005* (0.003)	-0.031 (0.025)
Training paid by employer	0.008*** (0.003)	0.025*** (0.009)	0.005* (0.002)	-0.038*** (0.014)
Training paid by employee	0.009 (0.008)	0.014 (0.019)	0 (0.003)	-0.023 (0.022)

Notes: This table shows AMEs computed on an ordered probit regression similar to specification (2) in Table 5.2 that includes a categorical variable that distinguishes either between training during and outside regular working hours or between training financed by the employer and the employee. Workers who did not participate in any training are the reference category in both cases. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 36,477.

Different types of informal learning

To analyse whether there are any differences in the relevance of different types of informal learning for workers' skill development, we distinguish between informal learning from others (co-workers and supervisors), informal learning by trial and error, or learning by doing, and informal learning by self-study.

Since the question for the different types of informal learning was only asked to those who reported a positive skill change (i.e. categories scoring six to 10), here we can use a sample of only 31,385 observations. The results in Table 5.8 show that, indeed, the relation between informal learning and skill improvement differs with the type of informal learning and it is heterogeneous across employees with different initial skill mismatch situations. We find, as shown in Panel A of Table 5.8, that, for workers in well-matching jobs, informal learning from colleagues and supervisors seems to contribute to their skill improvement as much as informal learning by trial and error and self-study. Underskilled workers, however, show greater skill improvement when learning by self-study than when learning from others, while there does not seem to be any significant difference in skill improvement between those who are involved in learning by trial and error and those who are not. In contrast, for the skill improvement of overskilled workers, informal learning from colleagues and supervisors seems to be as important as learning by doing, whereas learning by self-study does not make any significant contribution.²⁴ Panel B of Table 5.8 again shows that, in comparison with well-matched workers with similar informal learning participation, the relation between skill development and all three types of informal learning is stronger among workers with a job above their skill levels, and weaker for those with a job below their skill level.

5.5.2 On-the-job Learning Complementarity and Workers' Skill Development

In this section we raise the question whether or not the complementarity (or substitutability) between training and informal learning at work is related to the development of workers' skills.

First, we examine whether there is complementarity or substitution between training and informal learning. The estimation results of a multivariate analysis show that, as expected, training and informal learning on the job complement each other (see Table 5.9).²⁵ This complementarity holds for all well-matched, underskilled, and overskilled workers, although the strength of this complementarity differs between the three skill-mismatch groups (see Tables 5.A6 and 5.A7 in the Appendix). In comparison with initially well-matched workers, the complementarity is stronger among those who were initially underskilled for their job and slightly weaker but still positive among initially overskilled employees.

²⁴ Note, however, that we cannot make any inference regarding skill maintenance or decline due to sample truncation.

²⁵ The complementarity means that, on the one hand, the frequency of informal learning increases when a worker participates in training and, on the other hand, the average individual probability of training participation is higher the more often the worker engages in informal learning.

Table 5.8: AMEs of different types of informal learning on workers' skill improvement

Skill change	6	7	8	9	10
<i>A. AMEs within job–skill (mis)match groups</i>					
Initially well matched					
IL from others	-0.009*** (0.003)	-0.011*** (0.004)	-0.004*** (0.001)	0.006*** (0.002)	0.018*** (0.006)
IL by trial and error	-0.008*** (0.003)	-0.010*** (0.004)	-0.003*** (0.001)	0.005** (0.002)	0.016*** (0.006)
IL by self-study	-0.010*** (0.004)	-0.011*** (0.004)	-0.004*** (0.001)	0.008*** (0.002)	0.017*** (0.006)
Initially underskilled					
IL from others	-0.008*** (0.003)	-0.013*** (0.004)	-0.008*** (0.002)	0.004*** (0.001)	0.025*** (0.008)
IL by trial and error	-0.002 (0.002)	-0.003 (0.004)	-0.002 (0.002)	0.001 (0.001)	0.006 (0.007)
IL by self-study	-0.013*** (0.003)	-0.022*** (0.004)	-0.014*** (0.003)	0.007*** (0.002)	0.042*** (0.008)
Initially overskilled					
IL from others	-0.011** (0.005)	-0.011** (0.005)	-0.001** (0.001)	0.007** (0.003)	0.016** (0.007)
IL by trial and error	-0.011** (0.005)	-0.011** (0.005)	-0.002** (0.001)	0.007** (0.003)	0.017** (0.007)
IL by self-study	-0.002 (0.004)	-0.002 (0.005)	0.000 (0.001)	0.001 (0.003)	0.003 (0.007)
<i>B. AMEs between job–skill (mis)match groups (well matched ref.)</i>					
Initially underskilled					
IL from others	-0.022*** (0.002)	-0.028*** (0.003)	-0.020*** (0.002)	0.016*** (0.002)	0.054*** (0.006)
IL by trial and error	-0.029*** (0.004)	-0.038*** (0.005)	-0.015*** (0.003)	0.017*** (0.003)	0.065*** (0.009)
IL by self-study	-0.032*** (0.003)	-0.047*** (0.004)	-0.024*** (0.002)	0.018*** (0.002)	0.085*** (0.007)
Initially overskilled					
IL from others	0.007*** (0.003)	0.011*** (0.004)	0.003*** (0.001)	-0.009*** (0.002)	-0.012*** (0.004)
IL by trial and error	0.011*** (0.003)	0.013*** (0.004)	0.003*** (0.001)	-0.007*** (0.002)	-0.020*** (0.006)
IL by self-study	0.013*** (0.004)	0.015*** (0.005)	0.004*** (0.001)	-0.008*** (0.003)	-0.024*** (0.007)

Notes: This table shows AMEs computed on an ordered probit regression similar to specification (2) in Table 5.2 that includes three dummy variables to account for the three different types of informal learning. Since the question for the different types of informal learning was only asked to those who reported a positive skill change, the dependent variable skill change in this regression only takes values from six to 10. The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 31,385$.

Table 5.9: Complementarity between training and informal learning (AMEs)

Training		Probit AMEs			
<i>(IL never ref.)</i>					
	IL sometimes	0.178***			
		(0.019)			
	IL usually	0.239***			
		(0.019)			
	IL always	0.255***			
		(0.023)			
Informal learning		Never	Sometimes	Usually	Always
Oprobit AMEs					
Training		-0.019***	-0.072***	0.026***	0.065***
		(0.003)	(0.006)	(0.003)	(0.006)

Notes: All other controls are included. AMEs for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N= 37,187.

Second, to find out whether this complementarity plays a role in workers' skill development, we introduce interaction terms between the work-related learning variables in our main model and estimate the corresponding AMEs of training and informal learning conditionally evaluated on each other. This allows us to assess if the investment in each type of learning is related to any additional increase in the probability of skill improvement when the worker also engages in the other type of learning and if the probability of skill improvement increases further when the worker engages in both forms of learning.

Panel A of Table 5.10 shows that this is indeed the case. For example, on average, the probability of a high improvement of skills associated with training participation increases from 6.9 to 9.5 percentage points if the worker always participates in informal learning instead of never doing so.²⁶ The same holds for the skill development likelihood related to informal learning; it is significantly higher if workers also participate in training. This result suggests that training and informal learning provide workers with complementary skills rather than substitutable skills. Moreover, this complementarity appears to favour their skill accumulation.

The interaction outcome shown in Panel B of Table 5.10 indicates that the probability of a high improvement of skills when employees engage in both types of work-related learning (0.250) is significantly higher than the sum (0.224) of the partial contributions of training (0.069) and informal learning (0.155) if workers participate in only one of these forms of learning.²⁷ Consistent with Cunha and Heckman (2007) notions of direct complementarity and cross-productivity, the complementarity between training and informal learning is positively related to

²⁶ This difference (0.026) is significant with 99 percent confidence ($Chi^2(1) = 51.3, p = 0.000$). This is equivalent to the statistical difference (0.026) between the probability of a high improvement of skills associated with informal learning 'always' if workers also participate in training (0.181) and if they do not participate in training (0.155). This latter difference is also significant with 99 percent confidence ($Chi^2(1) = 35.9, p = 0.000$).

²⁷ This difference (0.026) is significant with 99 percent confidence ($Chi^2(1) = 127.1, p = 0.000$) and equal to the corresponding difference between the conditional contributions of training and informal learning, as explained in footnote 25.

a significant additional improvement of workers' skills. This suggests that not only training participation encourages informal learning and vice versa, but also the skills acquired through both forms of work-related learning seem to boost each other in a multiplier process that further reinforces the skill development of workers who participate in both types of work-related learning.

Table 5.10: Conditional AMEs of training and informal learning on workers' skill development

Skill change	0-4	5-6	7-8	9-10
<i>A. Conditional AMEs</i>				
AME of Training				
at IL never	-0.018*** (0.003)	-0.051*** (0.004)	0.000 (0.004)	0.069*** (0.005)
at IL sometimes	-0.016*** (0.001)	-0.053*** (0.004)	-0.012*** (0.002)	0.081*** (0.006)
at IL usually	-0.012*** (0.001)	-0.051*** (0.004)	-0.025*** (0.003)	0.088*** (0.006)
at IL always	-0.009*** (0.001)	-0.044*** (0.004)	-0.042*** (0.004)	0.095*** (0.007)
AME of IL sometimes				
at Training=0	-0.011** (0.005)	-0.024** (0.011)	0.012** (0.005)	0.023* (0.011)
at Training=1	-0.009** (0.004)	-0.026** (0.011)	0.000 (0.005)	0.035** (0.015)
AME of IL usually				
at Training=0	-0.023*** (0.006)	-0.056*** (0.012)	0.005 (0.004)	0.074*** (0.011)
at Training=1	-0.017*** (0.004)	-0.056*** (0.012)	-0.020*** (0.005)	0.093*** (0.012)
AME of IL always				
at Training=0	-0.034*** (0.006)	-0.104*** (0.013)	-0.017*** (0.005)	0.155*** (0.012)
at Training=1	-0.025*** (0.004)	-0.097*** (0.012)	-0.059*** (0.007)	0.181*** (0.014)
<i>2. Interaction effect (Training = 0 and IL never ref.)</i>				
at Training = 1 and IL always	-0.043*** (0.005)	-0.148*** (0.014)	-0.059*** (0.006)	0.250*** (0.018)

Notes: This table shows the conditional AMEs computed on an ordered probit regression similar to specification (2) in Table 5.2 that includes an interaction term between the training and informal learning variables. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 37,285.

5.5.2.1 Heterogeneity in Learning Complementarity and Workers' Skill Development

To assess if the additional improvement of skills associated with the complementarity between training and informal learning holds to the same extent for initially well-matched, underskilled, and overskilled employees, we estimate the AMEs of a three-way interaction term between the two forms of learning and the indicator for the initial skill mismatch in our main model. In Table 5.11 we present the most relevant estimation results.²⁸ The table shows that the com-

²⁸ Complete tables are available upon request.

plementarity between training and informal learning relates to additional gains in skill development only for initially well-matched²⁹ and overskilled³⁰ workers. Although training and informal learning are both associated with greater improvement in the skills of underskilled workers, the complementarity between these two types of learning does not seem to add any further to that skill improvement.³¹

If we extend the analysis of Cunha and Heckman (2007) to the workplace, as presented in Section 5.5.2, our findings are consistent with their notion of the skills multiplier. In their model, more complementarity of human capital implies greater productivity of new investments but, as the degree of complementarity increases, the role of the skill multiplier decreases. According to our findings, this holds for all three job–skill (mis)match groups in our analysis. Together, the results of Sections 5.5.1 and 5.5.2 suggest that having a higher initial stock of skills than required for the job implies a greater skill multiplier potential. However, this potential seems to be hampered by the tendency of overskilled workers to participate much less in both training and informal learning, in comparison to workers in a well-matching job. This situation not only seems to diminish the complementarity of human capital among overskilled workers, but also relates to a lower probability of developing their skills.

Conversely, in comparison to well-matched workers, those who were initially underskilled show a higher complementarity of human capital as they participate more often in training and informal learning, which seems to compensate for their disadvantage in the skill multiplier potential. This suggests that work-related learning helps in reducing the initial skill gap of underskilled workers. However, as implied by Heckman and Carneiro (2003), this learning could be more costly due to (1) the greater need for complementary investments in both training and informal learning to compensate for their initial disadvantage and (2) our finding that their lower initial stock of job skills seems to weaken the multiplier potential of these investments for their skill development. Nevertheless, initially underskilled workers still show the greatest improvement in skills among all workers who invest in their human capital while on the job.

²⁹ For workers in an initially well-matching job, the probability of a high improvement of skills when they engage in both types of work-related learning (0.253) is higher than the sum (0.232) of the partial contributions of training (0.073) and informal learning (0.159). The difference (0.021) is significant with 99 percent confidence ($Chi^2(1) = 96.7, p = 0.000$).

³⁰ For initially overskilled workers, the probability of a high improvement of skills when they engage in both types of work-related learning (0.289) is higher than the sum (0.255) of the partial contributions of training (0.070) and informal learning (0.187). The difference (0.032) is significant with 99 percent confidence ($Chi^2(1) = 159.6, p = 0.000$).

³¹ For initially underskilled workers, the probability of a high improvement of skills when they engage in both types of work-related learning (0.213) is statistically equivalent to the sum (0.212) of the partial contributions of training (0.079) and informal learning (0.133) ($Chi^2(1) = 0.4, p = 0.527$).

Table 5.11: Conditional AMEs of work-related learning on workers' skill development within and between job-skill s (mis)match groups

Skill change	0-4	5-6	7-8	9-10
Initially well matched				
<i>AME of Training</i>				
at IL never	-0.014*** (0.003)	-0.056*** (0.005)	-0.003 (0.005)	0.073*** (0.006)
at IL always	-0.007*** (0.001)	-0.045*** (0.004)	-0.042*** (0.004)	0.094*** (0.007)
<i>AME of IL always</i>				
at Training = 0	-0.026*** (0.006)	-0.108*** (0.016)	-0.025*** (0.004)	0.159*** (0.014)
at Training = 1	-0.019*** (0.004)	-0.097*** (0.015)	-0.064*** (0.007)	0.180*** (0.015)
<i>Interaction effect within well-matched workers (Training = 0 and IL never ref.)</i>				
Training = 1 and IL always	-0.033*** (0.006)	-0.153*** (0.017)	-0.067*** (0.005)	0.253*** (0.021)
Initially underskilled				
<i>AME of Training</i>				
at IL never	-0.007*** (0.002)	-0.039*** (0.008)	-0.033*** (0.007)	0.079*** (0.014)
at IL always	-0.003*** (0.001)	-0.028*** (0.005)	-0.049*** (0.009)	0.080*** (0.014)
<i>AME of IL always</i>				
at Training = 0	-0.010*** (0.003)	-0.062*** (0.018)	-0.061*** (0.013)	0.133*** (0.023)
at Training = 1	-0.006*** (0.002)	-0.051*** (0.015)	-0.077*** (0.018)	0.134*** (0.026)
<i>Interaction effect within underskilled workers (Training = 0 and IL never ref.)</i>				
Training = 1 and IL always	-0.013*** (0.004)	-0.090*** (0.018)	-0.110*** (0.015)	0.213*** (0.031)
<i>Interaction between underskilled and well-matched workers (Training=1 and IL always ref.)</i>				
Training = 1 and IL always	-0.005*** (0.001)	-0.036*** (0.004)	-0.057*** (0.008)	0.098*** (0.012)
Initially overskilled				
<i>AME of Training</i>				
at IL never	-0.036*** (0.006)	-0.066*** (0.006)	0.032*** (0.009)	0.070*** (0.005)
at IL always	-0.012*** (0.001)	-0.057*** (0.005)	-0.033*** (0.005)	0.102*** (0.009)
<i>AME of IL always at Training = 0</i>				
at Training = 1	-0.065*** (0.013)	-0.154*** (0.018)	0.032** (0.008)	0.187*** (0.015)
at Training = 1	-0.041*** (0.008)	-0.145*** (0.018)	-0.033*** (0.008)	0.219*** (0.017)
<i>Interaction effect within overskilled workers (Training = 0 and IL never ref.)</i>				
Training = 1 and IL always	-0.077*** (0.014)	-0.211*** (0.021)	-0.001 (0.014)	0.289*** (0.022)
<i>Interaction between overskilled and well-matched workers (Training=1 and IL always ref.)</i>				
Training = 1 and IL always	0.004*** (0.001)	0.028*** (0.006)	0.023*** (0.005)	-0.055*** (0.011)

Notes: This table shows conditional AMEs computed on an ordered probit regression similar to specification (2) in Table 5.3 that includes a third-way interaction term between the learning variables and the initial skill mismatch indicator. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 37,285.

5.6 Conclusions

In this chapter, we have analysed the extent to which training and informal learning on the job are related to the skill development of workers in 28 European countries. Consistent with assumptions from human capital theory, we find that employees who have been involved in training and informal learning show greater improvement of their skills. In line with Mincer's (1974) claim, we find that informal learning is more strongly related to the improvement of workers' skills than training participation is. However, this does not mean that participation in training is not important. Training and informal learning are shown to be complementary and this complementarity seems to further favour workers' skill development, since we find that those who participate in both forms of learning show a significant additional improvement of their skills. This is likely explained by the complementarity and cross-productivity between investments in human capital and their impact in the multiplier process of skills (Cunha and Heckman, 2007) that can also occur in the workplace, making the skills acquired by training and informal learning mutually-reinforcing.

We also analysed the heterogeneity in the relation between the two forms of job-related learning and workers' skill development with respect to their initial skill mismatch status. First, our results show that workers who participate in training or informal learning are more likely to improve their skills considerably than those with the same initial skill mismatch status who have not been involved in any learning activity. This result suggests that on-the-job learning helps improve the skills of both matched and mismatched workers. Second, compared to those who started in a job that matched well their skills, the skill development of underskilled workers appears to benefit the most from both training and informal learning, whereas the skill development of the overskilled benefits the least. In particular, human capital investments in the latter group seem to be more functional in offsetting skill depreciation and preserving their skill levels than in fostering skill accumulation.

The finding that underskilled employees who participate in training and informal learning show the largest increase in their skills suggests that human capital investments on the job could help to improve the job-skill match of these workers by closing the gap between their initial skill levels and the skills required in the workplace (Arulampalam et al., 2004). This could be motivated by a greater interest of these employees in maintaining their job or by richer learning opportunities offered at work (De Grip et al., 2008). In contrast, the finding that overskilled workers have the lowest probabilities of skill improvement and participation in both training and informal learning suggests that work-related human capital investments for these workers could be more relevant in maintaining the skills they do not use in their job or as a mitigating factor counteracting skill obsolescence (De Grip and Van Loo, 2002; Mincer and Ofek, 1982; Welch and Ureta, 2002), which could provide them with prospects to find a better job match in the current firm or other firms. It is, therefore, important to recognise the overskilled workers' need to participate in training and informal learning, even though they have the least increase in skills.

In our study, we find further heterogeneity across workers with a different

job–skill (mis)matches regarding the (1) complementary of human capital investments and (2) the multiplier potential of these investments for their skill development. On the one hand, overskilled workers are shown to have a greater skill multiplier potential but, at the same time, significantly lower rates of participation in both training and informal learning and lesser complementarity between these human capital investments than employees in a well-matching job. Accordingly, initially overskilled workers show the lowest probabilities of skill improvement. On the other hand, underskilled workers are more likely to participate in both training and informal learning and show greater complementarities between these human capital investments, which seems to compensate for the disadvantage in the skills multiplier potential, given their lower stock of job skills in comparison with well-matched employees. Underskilled workers therefore show the greatest probabilities of skill improvement among all workers who participate in work-related learning.

Our study then suggests that the interaction between work-related human capital investments and the utilisation of workers' initial stock of skills in their job plays an important role in the process of skill development. According to the literature on skills formation (e.g. Cunha and Heckman, 2007) as well as the literature on skill obsolescence (e.g. De Grip and Van Loo, 2002; Fernandez, 2002; Mincer and Ofek, 1982), being overskilled at job entry could be a negative context, not only because a worker's initial stock of skills may deteriorate due to non-use, but also because this deterioration may decrease the productivity and complementarity of further learning investments in the process of human capital accumulation. In contrast, being underskilled could be a positive context to raise the level and productivity of further learning investments. However, as implied by Heckman and Carneiro (2003), the elimination of the skill gap of these workers might be more costly due to the greater need for complementary investments in both training and informal learning and because their lower initial stock of job skills seems to weaken the multiplier potential of these investments in comparison with workers whose skills better match their job requirements.

Since lifelong learning and workers' skill development are essential for the productivity of firms with rapidly changing skill demands and for economic progress (World Economic Forum, 2014), knowledge about heterogeneities in the role of training and informal learning in workers' skill development with respect to their initial skill mismatch is crucial to make efficient decisions on human capital investments. Optimal decisions on investment in lifelong learning could contribute to reduce the misalignment between workers' potential productivity and the optimal productivity of their jobs. In this respect, it is important that more - in particular longitudinal - research be done on the causal effects of both formal training and informal learning on workers' skill development taking into account the basis and extent of the skill (mis)match they face in their jobs.

Appendices to Chapter 5

Table 5.A1: Distribution of the sample

<i>Country</i>	<i>Obs.</i>	<i>% Sample</i>	<i>% Initially well matched</i>	<i>% Initially underskilled</i>	<i>% Initially overskilled</i>
United Kingdom	2,830	7.6	41.8	24.0	34.2
Greece	1,450	3.9	41.9	19.8	38.3
Slovakia	835	2.2	41.9	36.1	22.0
Ireland	749	2.0	42.7	26.7	30.6
Austria	726	2.0	43.5	23.0	33.5
Finland	1,578	4.2	43.9	29.0	27.2
Cyprus	396	1.1	46.0	29.3	24.8
Czech Republic	1,278	3.4	48.4	32.9	18.7
Estonia	852	2.3	48.5	40.9	10.7
Lithuania	820	2.2	49.9	39.2	11.0
France	3,090	8.3	50.7	23.9	25.4
Poland	3,175	8.5	51.1	21.5	27.3
Spain	2,915	7.8	51.2	17.7	31.2
Germany	2,937	7.9	51.8	19.1	29.2
Latvia	808	2.2	52.4	37.0	10.6
Denmark	694	1.9	52.5	24.1	23.5
Belgium	1,003	2.7	52.6	20.2	27.1
Italy	2,283	6.1	53.8	20.5	25.8
Hungary	1,275	3.4	54.6	21.8	23.6
Bulgaria	882	2.4	55.7	27.1	17.2
Croatia	884	2.4	56.9	22.4	20.7
Netherlands	820	2.2	57.1	20.2	22.7
Sweden	742	2.0	57.3	19.3	23.5
Portugal	1,276	3.4	57.6	24.1	18.3
Malta	407	1.1	57.7	28.8	13.5
Romania	1,303	3.5	59.6	25.6	14.8
Slovenia	856	2.3	60.5	18.5	21.0
Luxembourg	421	1.1	73.6	11.4	15.0
TOTAL	37,285	100	50.94	23.92	25.14

Table 5.A2: Descriptive statistics

	<i>All</i>	<i>Initially well matched (51%)</i>	<i>Initially underskilled (24%)</i>	<i>Initially overskilled (25%)</i>
Training (during tenure)	0.62	0.61	0.70	0.58
Training 12 months	0.57	0.56	0.60	0.55
Training during working hours	0.45	0.43	0.49	0.43
Training outside working hours	0.22	0.21	0.19	0.24
Training paid by the employer	0.69	0.69	0.71	0.66
Training paid by the employee	0.12	0.11	0.10	0.16
IL never	0.04	0.04	0.02	0.05
IL sometimes	0.41	0.40	0.38	0.45
IL usually	0.33	0.34	0.36	0.30
IL always	0.22	0.22	0.23	0.20
IL from others*	0.77	0.76	0.86	0.72
IL by trial and error*	0.61	0.58	0.70	0.58
IL by self-study*	0.56	0.52	0.63	0.55
Formal education (during tenure)	0.14	0.13	0.17	0.11
<i>Individual characteristics</i>				
Age (24-65) s.d. = 9.8	42.1	42.39	41.33	42.25
Female	0.39	0.39	0.42	0.37
Low level of education	0.12	0.13	0.12	0.10
Intermediate level of education	0.41	0.43	0.42	0.38
High level of education	0.47	0.44	0.47	0.52
Years of tenure (0-41) s.d.= 9.3	10.50	10.88	11.33	8.93
Permanent contract	0.87	0.87	0.88	0.85
Fixed temporary contract	0.1	0.10	0.09	0.12
Temporary agency contract	0.01	0.01	0.01	0.01
No formal contract	0.02	0.02	0.02	0.02
Telephone (interviewed)	0.21	0.23	0.24	0.13
<i>Industry</i>				
Agriculture	0.02	0.02	0.02	0.02
Manufacturing	0.19	0.19	0.21	0.18
Construction	0.06	0.07	0.06	0.05
Sales and transportation	0.20	0.19	0.17	0.23
Information and communication	0.07	0.07	0.08	0.07
Financial and real state	0.06	0.06	0.06	0.06
Professional and Tech	0.07	0.07	0.08	0.06
Public administration	0.25	0.26	0.25	0.25
Other services	0.08	0.08	0.07	0.08
<i>Occupation</i>				
Managers	0.09	0.08	0.09	0.10
Professionals	0.22	0.22	0.24	0.18
Technicians	0.17	0.17	0.19	0.15
Service and sales workers	0.12	0.12	0.11	0.14
Clerical support	0.21	0.2	0.18	0.24
Skilled agricultural	0.01	0.01	0.01	0.01
Building, crafts or related trades	0.08	0.09	0.09	0.06
Plant and machine operators	0.07	0.07	0.07	0.08
Elementary	0.04	0.04	0.03	0.05
<i>Firm size</i>				
1-9	0.20	0.20	0.20	0.20
10-49	0.28	0.28	0.29	0.27
50-99	0.13	0.14	0.12	0.13
100-249	0.13	0.13	0.13	0.14
250-499	0.08	0.08	0.08	0.09
>500	0.17	0.17	0.18	0.17
<i>Observations</i>	37,285	18,992	8,919	9,374

*For these variables we have fewer observations (31,954). It is because the respective questions were only asked to respondents who reported a positive skill change in our dependent variable (i.e. above category 5 in the 0-10 scale).

Table 5.A3: Ordered probit coefficients for skill development

Skill change	(1)	(2)
Training	0.255*** (0.022)	0.251*** (0.022)
IL sometimes	0.145** (0.056)	0.122* (0.064)
IL usually	0.300*** (0.061)	0.258*** (0.069)
IL always	0.540*** (0.067)	0.501*** (0.073)
Formal education	0.154*** (0.021)	0.139*** (0.031)
Initially underskilled	0.304*** (0.016)	0.461*** (0.084)
Initially overskilled	-0.218*** (0.026)	-0.389*** (0.081)
Training # Initially underskilled		-0.035 (0.036)
Training # Initially overskilled		0.039* (0.021)
IL sometimes # Initially underskilled		-0.160** (0.077)
IL sometimes # Initially overskilled		0.114 (0.080)
IL usually # Initially underskilled		-0.151** (0.075)
IL usually # Initially overskilled		0.217*** (0.080)
IL always # Initially underskilled		-0.170** (0.081)
IL always # Initially overskilled		0.218** (0.091)
Formal education # Initially underskilled		-0.024 (0.033)
Formal education # Initially overskilled		0.103** (0.044)
Age	-0.007*** (0.001)	-0.007*** (0.001)
Female	0.209*** (0.018)	0.209*** (0.018)
Intermediate level of education	-0.077*** (0.025)	-0.073*** (0.025)
High level of education	-0.213*** (0.029)	-0.209*** (0.030)
Years of tenure	0.016*** (0.001)	0.016*** (0.001)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>
cut1	-2.598*** (0.084)	-2.635*** (0.088)
cut2	-2.256*** (0.076)	-2.292*** (0.081)
cut3	-1.939*** (0.073)	-1.975*** (0.078)
cut4	-1.658*** (0.071)	-1.694*** (0.076)
cut5	-1.392*** (0.070)	-1.427*** (0.076)
cut6	-0.484*** (0.069)	-0.517*** (0.075)
cut7	-0.142** (0.068)	-0.155** (0.075)
cut8	0.476*** (0.069)	0.444*** (0.074)
cut9	1.222*** (0.070)	1.191*** (0.075)
cut10	1.801*** (0.070)	1.769*** (0.075)
<i>Pseudo R2</i>	0.0701	0.0709
<i>BIC-stat (27)</i>	127,204.90	127,147.80
<i>LR chi2 (10)</i>		37.13 (0.0001)

Notes: Specification (1) gives the coefficients of the regression without interacting the learning variables and the workers' skill mismatch status, and specification (2) includes these interactions. The dependent variable skill change is measured from categories zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). Other controls include dummies for changes in job-position and/or task content, the working process and the need of learning new things since the start of the job as well as dummies for type of contract, occupation, industry, firm size, country and survey answered by phone. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 37,285.

Table 5.A4: Estimations of training and informal learning participation

	(1) Probit AME Training	(2) Probit AME IL	(3) OLS IL intensity
Initially underskilled	0.068*** (0.007)	0.048*** (0.004)	0.055*** (0.012)
Initially overskilled	-0.014*** (0.005)	-0.016*** (0.003)	-0.094*** (0.017)
Age	-0.005** (0.002)	-0.001** (0.004)	-0.010** (0.004)
$Age^2 * 100$	0.007** (0.003)	0.003** (0.001)	0.007** (0.003)
Female	-0.007 (0.008)	0.001 (0.004)	-0.017 (0.016)
Intermediate level of education	0.051*** (0.012)	0.008 (0.006)	0.077** (0.036)
High level of education	0.089*** (0.013)	0.020*** (0.006)	0.124*** (0.036)
Years of tenure	0.009*** 0.000	-0.001*** 0.000	-0.002** (0.001)
Temporary contract	-0.083*** (0.009)	0.013* (0.008)	0.082*** (0.016)
Agency contract	-0.154*** (0.039)	0.034*** (0.012)	0.122** (0.056)
No formal contract	-0.166*** (0.023)	0.001 (0.008)	0.035 (0.046)
Learning attitude (std)	0.022*** (0.003)	-0.002 (0.002)	0.147*** (0.015)
<i>Other controls</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>

Notes: Columns (1) and (2) in this table show AMEs computed based on probit regressions. Column (3) reports OLS coefficients. Other controls include occupation, industry, firm size and country dummies. The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N=37,285.

Table 5.A5: AMEs of work-related learning on workers' skill development

Skill change	0-4	5-6	7-8	9-10
<i>A. All employees</i>				
Training	-0.019*** (0.002)	-0.046*** (0.003)	-0.007*** (0.001)	0.072*** (0.006)
IL (dummy)	-0.015*** (0.003)	-0.053*** (0.010)	-0.021*** (0.007)	0.091*** (0.013)
<i>B. AMEs within the same initial job-skill (mis)match group</i>				
Initially well-matched				
Training	-0.013*** (0.002)	-0.046*** (0.006)	-0.014*** (0.002)	0.073*** (0.007)
IL (dummy)	-0.013*** (0.001)	-0.054*** (0.004)	-0.023*** (0.002)	0.090*** (0.008)
Initially underskilled				
Training	-0.007*** (0.001)	-0.043*** (0.007)	-0.018*** (0.005)	0.068*** (0.012)
IL (dummy)	-0.014** (0.006)	-0.038** (0.016)	-0.004* (0.002)	0.056** (0.025)
Initially overskilled				
Training	-0.018*** (0.002)	-0.054*** (0.004)	-0.005** (0.002)	0.077*** (0.006)
IL (dummy)	-0.043*** (0.006)	-0.064*** (0.009)	0.008 (0.006)	0.099*** (0.013)

Notes: This table shows the AMEs computed on an ordered probit specification similar to column (2) in Table 5.A3 that includes a dummy variable for IL instead of the original categorical variable. The dependent variable skill change is measured by 11 ordinal categories from zero to 10 (0= skills have worsened a lot, 5= skills have stayed the same, 10= skills have improved a lot). AMEs on skill change are grouped into four categories: worsened (0-4), no or hardly any change (5-6), intermediate improvement (7-8), and high improvement (9-10). The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 37,285$.

Table 5.A6: Training (probit) estimations including informal learning as explanatory variable

Training	Probit AMEs
A. Initially well matched	
IL sometimes	0.173*** (0.022)
IL usually	0.233*** (0.024)
IL always	0.259*** (0.030)
<i>B. AMEs between the initial job-skill (mis)match groups (well matched ref.)</i>	
Initially underskilled	
IL sometimes	0.074*** (0.011)
IL usually	0.064*** (0.009)
IL always	0.046*** (0.010)
Initially overskilled	
IL sometimes	-0.010 (0.007)
IL usually	0.006 (0.010)
IL always	-0.017 (0.012)

Notes: All other controls are included. The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N= 37,187.

Table 5.A7: Informal learning (oprobit) estimations including training as explanatory variable

Informal learning	Never	Sometimes	Usually	Always
A. Initially well matched				
Training	-0.017*** (0.003)	-0.072*** (0.008)	0.024*** (0.003)	0.065*** (0.007)
<i>B. AME between the initial job-skill (mis)match groups (well matched ref.)</i>				
Initially underskilled				
Training	-0.002 (0.001)	-0.008 (0.006)	0.002 (0.001)	0.007 (0.006)
Initially overskilled				
Training	0.005*** (0.001)	0.026*** (0.007)	-0.006*** (0.002)	-0.024*** (0.007)

Notes: All other controls are included. The AME for categorical variables is the discrete change from the base level. Standard errors clustered at country level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N= 37,187.

Figure 5.A1: AMEs of work-related learning on workers' skill development (cat. 9-10) within job-skill (mis)match groups over years of tenure.

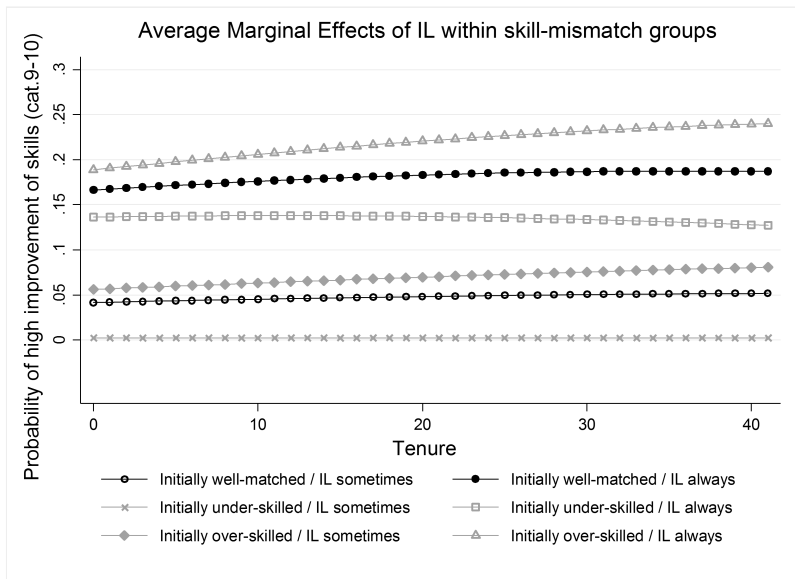
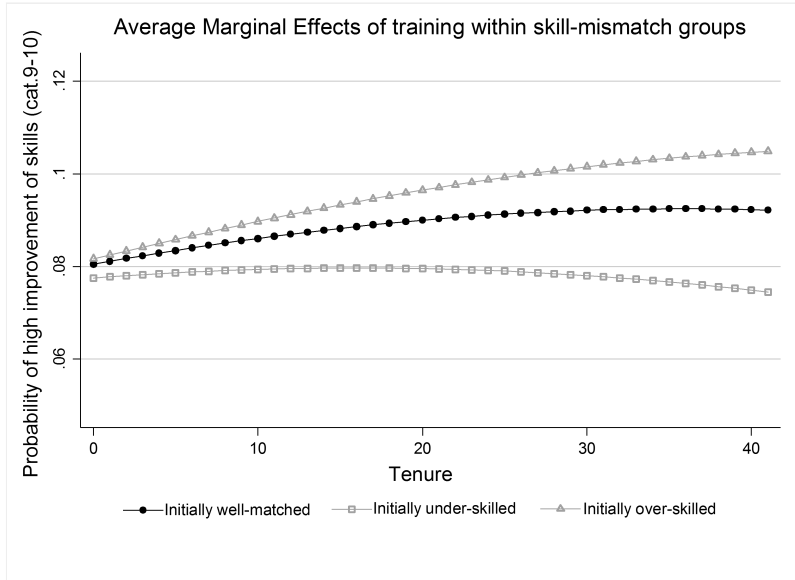
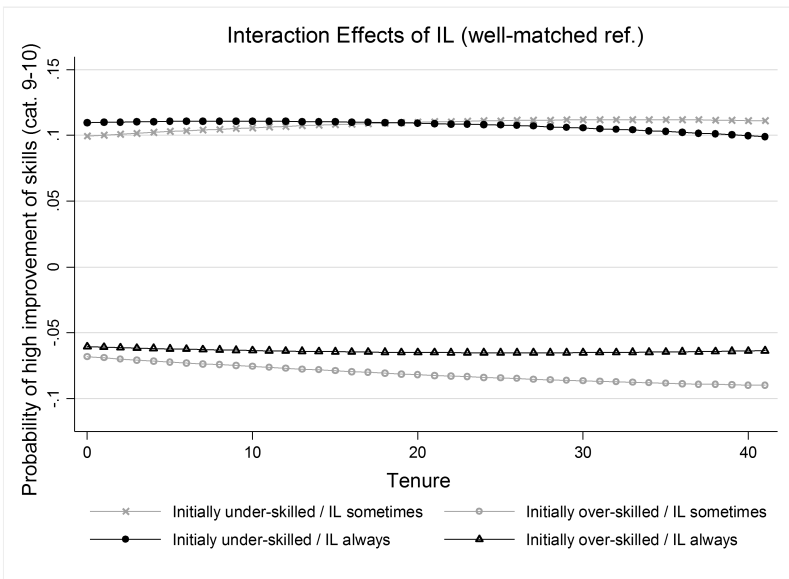
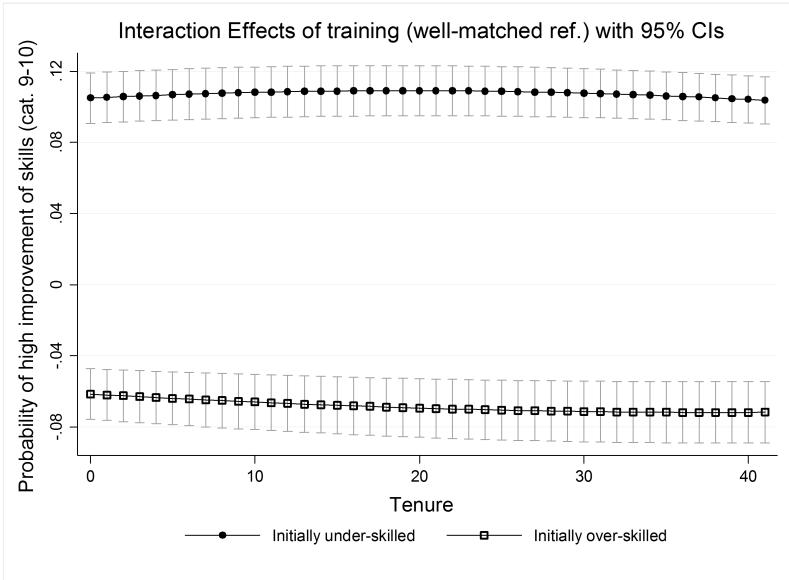


Figure 5.A2: AMEs of work-related learning on workers' skill development (cat. 9-10) between job-skill (mis)match groups over years of tenure.



Conclusion

The studies presented in this thesis contribute to the literature on human capital development both in schools and the workplace. First, Chapters 2 and 3 add to the growing research on the effects of grade retention on children's further skill development. These two chapters contribute to a better understanding and measurement of the positive and negative direct effects, as well as the spill-over effects to classmates, of school failure on school performance and school dropout. Second, Chapters 4 and 5 add new insights into the determinants of human capital accumulation in the workplace via both formal and informal learning, and into the relation between both modes of learning and workers' skill development, mismatch and depreciation. In the following paragraphs, we summarise the main findings of these four chapters.

In Chapter 2, we analysed the effects of grade retention on students' performance in their final math and language tests at the end of secondary education. By exploiting plausibly exogenous variation in retention probabilities from a schools' retention reform in Colombia, we find no significant effects on math scores that can be attributed either to retained or non-retained pupils. In contrast, we do find a positive effect of increased retention on the language performance of the retained students, especially those at the lower end of the distribution. For non-retained students, we find the effect of increased retention to be negative on their language scores. This effect is shown to be driven by the low-performing students as well.

The main contribution of this study is the analysis of the non-linear effect of retention on test scores at various moments of the ability distribution. We show that modest increases in retention lead to higher scores in language for the retained students, but when many students are retained, such gains decrease. These results imply that retention has marginal decreasing performance returns and, therefore, it is important for schools to retain students at optimal levels. Finally, we provide evidence that neither average class size nor teachers' average educational achievements are the mechanisms that explain our main findings. In contrast, recruiting more high-quality new teachers in the school seems to be the channel through which increased retention affects school performance.

In Chapter 3, we estimated the effects of grade retention on school dropout rates in secondary schools, distinguishing between (1) the end-of-the-year effect for retained students and (2) the effect in the year after retention for all pupils

enrolled in each grade. We analysed the heterogeneity of these effects depending on the timing at which retention occurs by estimating the impact of retention on dropouts at each grade of secondary education from grade 6 to grade 11.

By exploiting the same policy reform as in Chapter 2, we highlight two major findings. First, a remarkably large positive effect of grade retention on end-of-year dropout rates among retained students and a positive, although relatively small, effect of grade failure on consecutive dropout rates among all students enrolled in the year following retention. Second, we find significant heterogeneity in both effects depending on the timing of retention along the secondary education cycle: the effects of grade failure on early dropout rates are stronger when retention takes place at the earlier grades whereas the effect on retained students is strongest if retention occurs at grade 9 and grade 11. These are precisely the grades where successful completion entitles the students to receive the lower secondary school certificate and the high-school diploma, respectively.

Furthermore, we find empirical evidence of the heterogeneity and non-linearity of the main effects with regard to the treatment intensity. We observe that retained students from grade 9 to 11 in mid-treated schools had greater dropout rates than similar students in high-treated schools where retention rates were even higher. This last finding suggests that students retained at later stages of secondary education in a school environment where retention is not a common outcome can cause more harm to the future perspectives of those students than if they would have been in schools where retention was more pronounced or a more likely result at the end of the school year.

In Chapter 4, we analysed the difference in informal learning at work between temporary and permanent male employees across 20 OECD countries. Remarkably, we find that workers in temporary jobs engage more intensively in informal learning than permanent employees do, although the former are, indeed, less likely to participate in formal training. These results account for the endogeneity of the selection into temporary contracts by estimating FIML endogenous switching regression models that exploit workers' differential exposure to unemployment and employment protection legislation across countries and age groups.

Our findings in Chapter 4 could be driven by several mechanisms. First, the hypothesis that temporary workers could substitute informal learning for the lack of formal training is not supported by our analysis. Instead, we find complementarity between these two types of learning. Second, we do not find support for the view that high-skilled workers drive the difference in informal learning between temporary and permanent employees. However, we find descriptive evidence suggesting that the higher informal learning investments of temporary employees might be driven by the incentives of finding a more stable job. We find that the positive influence of temporary contracts is larger earlier in workers' careers and that the intensity of informal learning is likely higher for temporary workers who have better expectations of job stability. Our finding suggests that those who know about their better chances of obtaining a permanent job position, given their own characteristics, are those expected to show a higher intensity of informal learning at work, which could help to improve these

chances in their favour.

Finally, in Chapter 5, we analysed the extent to which training and informal learning on the job are related to the skill development of workers in 28 European countries. In line with assumptions from human capital theory, we find that employees who have been involved in training and informal learning show greater improvement of their skills. We also find that training and informal learning are complementary, and this complementarity seems to further favour workers' skill development.

In this chapter, we also analysed the heterogeneity in the relation between the two forms of job-related learning and workers' skill development with respect to their initial skill mismatch status. First, our results show that workers who participate in training or informal learning are more likely to improve their skills considerably than those with the same initial skill mismatch status who have not been involved in any learning activity. Second, compared to those who started in a job that matched well their skills, the skill development of underskilled workers appears to benefit the most from both training and informal learning, whereas the skill development of the overskilled benefits the least. These findings suggest that the interaction between work-related human capital investments and the utilisation of workers' initial stock of skills in their job plays an important role in the process of skill development. First, human capital investments on the job could help to improve the job-skill match of underskilled workers by closing the gap between their initial skill levels and the skills required in the workplace. Second, human capital investments among overskilled workers seem to be more relevant in maintaining the skills they do not use in their job or as a mitigating factor counteracting skill obsolescence than in fostering skill accumulation.

Since human capital development throughout the life cycle is a major societal challenge related to various issues ranging from pre-school education to lifelong learning, it launches a vast research agenda. This thesis attempted to contribute with new empirical insights into some of the many relevant research questions related to school retention and lifelong learning in the workplace. In the following Valorisation Addendum, we further go into the relevance of the outcomes of this thesis for policy making and society in general.

Valorisation Addendum

The research results presented in this dissertation can be a source of evidence and inspiration for policy makers, schools, firms and workers when making decisions to foster skill development throughout the lifecycle. Nowadays, both in schools and the workplace, policy makers and other decision makers face several challenges in this field. Harnessing the best returns of skills requires a complex assessment of timely and well-matching investments in human capital formation related to initial education, the amount and quality of the abilities available in the population, the skills required in the labour market, and the development and effective use of those skills in good jobs.

Chapters 2 and 3 in this thesis focus on the effects of school grade retention on children's skill formation and development in the context of a developing country, where universalisation of education continues to be a challenge. Chapter 2 analyses the effects on school performance (math and language test scores) and Chapter 3, the effects on school dropout. On the one hand, Chapter 2 shows that higher retention in secondary school does not have any significant effect on the math performance of retained and non-retained students. On the other hand, increased retention is shown to have a modest positive impact on the language performance of retained students. This result suggests that by repeating a grade, students at the lower end of the ability distribution could gain a more thorough understanding of the material in certain subjects that enables them to perform better in later stages of their lives. Very importantly, these positive effects are nonlinear, which implies that increased retention has marginally decreasing returns on students' performance. Identifying the optimal levels of retention rates is one of the critical implications for schools, given that at some point in the lower end of the ability distribution, students will no longer benefit from being retained.

Making the decision of when lower performing students should be retained or promoted becomes more complex when the spill-over effects of retention on the non-retained peers are taken into consideration. According to our analyses, higher retention leads non-retained students to perform worse in language tests. This finding implies that the positive selection effects for non-retained students may be dominated by the plausibly negative influence of less able peers in the classroom because, for instance, teachers could feel inclined to adjust the level of their teaching downward because of the higher share of repeating students in the classroom. Our study shows the importance of analysing the effects of retention simultaneously for retained and non-retained students and at different margins of the ability distribution to be able to estimate more accurately the net effects of retention at aggregated levels. A better estimation of these effects will better

inform policy makers and schools not only in their decisions on whether to retain or promote low performing students but also in their consideration of alternative policies to grade retention such as early class repetition or reinforcement and/or reallocation of students to more similar groups.

In Chapter 3, I further investigate the effects of grade retention on school dropout rates during secondary education. The findings of this chapter complement the analyses of the previous chapter by enriching our understanding of whether the performance gains and losses of retention can be outweighed by lower costs of school dropout or strengthened by higher costs of school dropout. In this chapter, we find a remarkably large positive effect of grade retention on end-of-year dropout rates among retained students and a positive, although relatively smaller, effect of grade failure on consecutive dropout rates among all students enrolled in the year following retention. This finding contributes to the evaluation of the costs and benefits of retention practices for society. It is important to take both the effects of retention for retained and non-retained students into consideration, analysing both the positive and negative effects retention may have not only on school performance but also on the decisions of continuing secondary school studies until graduation. First, this study suggests that there are differences between the processes of dropping out early and dropping out late that would be hidden by considering all dropouts together. Second, the result that increased retention has a large impact on the probability of children leaving their studies questions whether the moderate gains of retention on children language performance are a sufficient argument for a higher retention rate. This could imply either a trade-off or a double challenge for policy makers and schools in developing countries, where skill development is crucial for social mobility but also where school dropout is a very serious issue given the limitations in realising the universalisation of schooling.

Moreover, in Chapter 3 we analyse the heterogeneous effects of retention on dropouts depending on the timing of the retention. We have distinguished retention from grade 6 to grade 11 of secondary education. Notably, we find that the overall positive effect of retention on dropout rates for the full cohort of students on the year following retention is stronger in the earlier grades of secondary school and that there is not any significant effect during the last two years of high school. These results imply that the strongest dropout effect takes place during the year of transition from primary to secondary education, i.e. grade 6. This also implies, in light of the human capital literature, that earlier investments to remedy skill deficits might be much more effective than later attempts to repair such deficits. Conversely, the retention effect on the dropout rates of retained students at the end of the year of their retention is strongest if students are retained at the end of grade 9 and grade 11, suggesting that particular attention to retained students at these two specific grades is necessary since leaving school at these stages would imply the highest costs of retention due to the forgone opportunity for students to finalise either one of the two cycles of secondary education and receive the corresponding certificate.

The studies presented in Chapters 2 and 3 emphasise the importance of developing a more holistic approach to assess the effects of retention. To answer the question of whether – and when – low-achieving students should repeat

a failed grade or not, policy makers and school directors need to take into consideration more than the extent to which grade retention affects the retained individuals. It should also be considered to what extent grade repeaters may affect their classmates when they repeat the grade in the following year as well as whether the context and the timing mitigate the undesirable effects of grade retention or intensify its positive effects on the human capital development of pupils. More accurate procedures to identify the risks and benefits of grade retention will help schools' awareness of the extent to which they need to retain students or provide interventions of a different kind instead.

In Chapter 4, we study the difference in informal learning at work between temporary and permanent workers across 20 OECD countries. Remarkably, we find that workers in temporary jobs engage more intensively in informal learning than employees with a permanent contract do, although the former are, indeed, less likely to participate in formal training. However, the hypothesis that temporary workers could substitute informal learning for the lack of formal training is not supported by our analysis. Instead, we find complementarity between these two types of learning, which is consistent with the idea that human capital accumulated through different sources not only enhances skills but also increase the learning capacity in a dynamic complementary process. The results of this study suggest that temporary jobs could incentivise higher investments in informal human capital for workers to increase their chances of promotion to a more secure job. This finding implies that workers perceive more intense learning as a profitable investment in their career development, particularly when facing the uncertainty of a temporary contract.

Chapter 4 derives several implications. First, it highlights the importance of informal learning for human capital development and that learning is a life-time matter that also takes place outside of schools and continues throughout life. Skill development through informal learning in the workplace, or work experience, is often neglected in policy discussions because of the difficulties to measure it or the current emphasis on formal education. However, the findings in this chapter suggest that informal learning in the workplace is an important source of career development in a modern economy. Once we recognise and make more visible the importance of informal sources of learning for skill formation and development, we could think about policies to foster skill in many other different ways.

Second, our analyses point towards a potential positive feature of temporary contracts that has important implications. Temporary jobs need not be dead-end jobs. If temporary jobs are taken by individuals in lieu of unemployment in search for further individual promotion in the labour market, these jobs could offer them opportunities for learning by doing particular tasks and productive interaction with other workers. This is important not only as a source of productive accumulation of human capital while working but also as a potential advantage for individuals who would otherwise be unemployed. Such jobs with high learning content could then be a stepping stone towards more stable employment.

Therefore, the challenge for firms and policy makers is in harnessing the learning potential of temporary employment. The findings of this chapter sug-

gest the existence of two different kinds of temporary jobs in terms of their learning opportunities: (1) good temporary jobs, with high levels of task autonomy and collaboration, offering good opportunities for training and informal learning and, likely, leading to positive career expectations of upward mobility, and (2) bad temporary jobs, which have no or very few opportunities to foster workers' human capital. In the latter jobs, workers could get trapped in a cycle between precarious jobs and unemployment. This implies labour segmentation within temporary employment due to the distinction between jobs of low and high learning content. Since policy makers have already stressed the importance of finding an appropriate balance between labour market flexibility and job security, policies should be supported by analysing how contract incentives influence workers' skill investments and their career development expectations. The design of these policies underlines the importance of improving firms' learning strategies to optimise the benefits of both training and informal learning to foster sustainable employability of a flexible workforce as well as the aggregate productive capacity of the economy in the long term.

In Chapter 5, we analyse the extent to which formal training and informal learning on the job are related to the skill development of workers across 28 European countries. The results confirm that investments in both types of on-the-job learning contribute to a greater improvement of workers' skills, although the contribution of informal learning to this improvement seems to be the strongest. However, this result does not suggest that participation in training is not important or could be substituted by informal learning. Our results suggest that investments in training and informal learning are complementary, which empirically validates a key feature of the technology of skill development in the sense that the skills acquired by training and informal learning can mutually reinforce each other in a multiplier process with strong synergistic components over the life cycle.

The results described above are, moreover, heterogeneous with respect to the workers' initial job-skill mismatch status. We find that, compared to those in a job that matches well with their skills, the skill development of under-skilled workers appears to benefit most from both training and informal learning, whereas the skill development of the overskilled benefits the least. This finding suggests that on-the-job learning helps to improve the skills of both matched and mismatched workers, nonetheless to a different extent. The finding for under-skilled employees suggests that human capital investments on the job could help to improve the job-skill match of these workers by closing the gap between workers' potential productivity and the productivity of the job in which they are employed. In contrast, the finding for overskilled employees suggests that work-related human capital investments for these workers play a more critical role in the restoration and replenishment of human capital in the context of skill depreciation, that is, human capital investment could be more relevant in maintaining the skills they do not use in their job or as a mitigating factor counteracting skill obsolescence, which could eventually provide them with prospects to find a better job match in the current firm or other firms.

This study then suggests that the interaction between work-related human capital investments and the utilisation of workers' initial stock of skills in their

job plays an important role in the process of skill development. Since lifelong learning and workers' skill development are essential for the productivity of firms as well as macro-economic growth, knowledge about heterogeneities in the role of training and informal learning in workers' skill development with respect to their initial skill mismatch is crucial to make efficient decisions on human capital investments. Optimal decisions on investment in lifelong learning could contribute to reduce the misalignment between workers' potential productivity and the optimal productivity of their jobs.

Overall, this thesis emphasises that developing and maintaining human capital throughout the lifecycle is a key challenge for several decision makers in both the fields of education and lifelong learning while at work. Skill development policies require coherence and cooperation across diverse areas and levels of governments, as well as with the private sector, schools, teachers and workers. Skills development could be more effective if the world of learning and the world of work are better linked. This holds in both developed and developing countries. In the end, earlier investing in human capital and good quality jobs is likely to be in the long run much less costly than remediation interventions to combat poorer health, lower incomes, unemployment and social exclusion as related outcomes of lower levels of human capital development.

Bibliography

- Abraham, K. G. (1990). Restructuring the employment relationship: The growth of market-mediated work arrangements. *New developments in the labor market: Toward a new institutional paradigm*, 85–119.
- Acemoglu, D. and J. S. Pischke (1999a). Beyond becker: Training in imperfect labour markets. *The Economic Journal* 109(453), 112–142.
- Acemoglu, D. and J.-S. Pischke (1999b). The structure of wages and investment in general training. *Journal of Political Economy* 107(3), 539–572.
- Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics letters* 80(1), 123–129.
- Alexander, K., E. D. and N. Kabbani (2001). The dropout process in life course perspective. *Teachers College Record* 70, 87–107.
- Allen, J. and R. Van der Velden (2001). Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers* 53(3), 434–452.
- Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics* 114(2), 533–575.
- Angrist, J. D. and J. S. Pischke (2008). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press.
- Arrow, K. J. (1962). The economic implications of learning by doing. *The review of economic studies* 29(3), 155–173.
- Arulampalam, W. and A. L. Booth (1998). Training and labour market flexibility: is there a trade-off? *British Journal of Industrial Relations* 36(4), 521–536.
- Arulampalam, W., A. L. Booth, and M. L. Bryan (2004). Training in europe. *Journal of the European Economic Association* 2(2-3), 346–360.
- Atkinson, A. (1998). Exclusion, employment and opportunity. Technical Report CASE Working paper 4, Centre for Analysis of Social Exclusion, London School of Economics.
- Autor, D. H. (2001). Why do temporary help firms provide free general skills training? *The Quarterly Journal of Economics* 116(4), 1409–1448.

- Barrett, A. and P. J. O’Connell (2001). Does training generally work? the returns to in-company training. *ILR Review* 54(3), 647–662.
- Bartel, A. P. (1994). Productivity gains from the implementation of employee training programs. *Industrial relations: a journal of economy and society* 33(4), 411–425.
- Bauernschuster, S., T. Hener, and H. Rainer (2016). Children of a (policy) revolution: The introduction of universal child care and its effect on fertility. *Journal of the European Economic Association* 14(4), 975–1005.
- Becker, G. S. (1964). *Human Capital: A theoretical and empirical analysis, with special reference to education*. National Bureau of Economic Research.
- Belot, M. and V. Vandenberghe (2014). Evaluating the “threat” effects of grade repetition: Exploiting the 2001 reform by the french-speaking community of Belgium. *Education Economics* 22(1), 73–89.
- Belzil, C. and M. Leonardi (2007). Can risk aversion explain schooling attainments? evidence from italy. *Labour Economics* 14(6), 957–970.
- Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75(4), 352–365.
- Berton, F. and P. Garibaldi (2012). Workers and firms sorting into temporary jobs. *The Economic Journal* 122(562), 125–154.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust Differences-in-Differences estimates? *The Quarterly Journal of Economics* 119(1), 249–275.
- Blázquez, M. and M. Jansen (2008). Search, mismatch and unemployment. *European Economic Review* 52(3), 498–526.
- Blundell, R., L. Dearden, C. Meghir, and B. Sianesi (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal studies* 20(1), 1–23.
- Booth, A. L. (1991). Job-related formal training: who receives it and what is it worth? *Oxford Bulletin of Economics and Statistics* 53(3), 281–294.
- Booth, A. L., M. Francesconi, and J. Frank (2002). Temporary jobs: stepping stones or dead ends? *The economic journal* 112(480).
- Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research* 103(3), 191–207.
- Bowers, A. J. and R. Sprott (2012). Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *The Journal of Educational Research* 105(3), 176–195.

- Bowers, A. J., R. Sprott, and S. A. Taff (2012). Do we know who will drop out? a review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal*, 77–100.
- Brutti, Z. and F. Sánchez (2017). Does better teacher selection lead to better students? evidence from a large scale reform in colombia. *Documento de trabajo CEDE* (11).
- Buccioli, A. and L. Zarri (2015). The shadow of the past: financial risk taking and negative life events. *Journal of Economic Psychology* 48, 1–16.
- Busso, M., M. Bassi, and J. S. Muñoz (2013). Is the glass half empty or half full? school enrollment, graduation, and dropout rates in latin america.
- Cabrol, M. (2002). Los desafíos de la educación secundaria:¿ que nos dice el análisis de flujos? Technical report, Inter-American Development Bank.
- Caliendo, M. and R. Schmidl (2016). Youth unemployment and active labor market policies in europe. *IZA Journal of Labor Policy* 5(1), 1.
- Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014). Violence and risk preference: experimental evidence from afghanistan. *The American Economic Review* 104(1), 123–148.
- Cameron, L. and M. Shah (2015). Risk-taking behavior in the wake of natural disasters. *Journal of Human Resources* 50(2), 484–515.
- Chevalier, A. (2003). Measuring over-education. *Economica* 70(279), 509–531.
- Chevalier, A. and J. Lindley (2009). Overeducation and the skills of uk graduates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(2), 307–337.
- Chuang, Y. and L. Schechter (2015). Stability of experimental and survey measures of risk, time, and social preferences: a review and some new results. *Journal of Development Economics* 117, 151–170.
- Cockx, B. and M. Picchio (2012). Are short-lived jobs stepping stones to long-lasting jobs? *Oxford Bulletin of Economics and Statistics* 74(5), 646–675.
- Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015). Evidence for countercyclical risk aversion: an experiment with financial professionals. *The American Economic Review* 105(2), 860–885.
- Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016). From late to mte: alternative methods for the evaluation of policy interventions. *Labour Economics* 41, 47–60.
- Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review* 97(2), 31–47.

- Cutuli, G. and R. Guetto (2012). Fixed-term contracts, economic conjuncture, and training opportunities: a comparative analysis across european labour markets. *European Sociological Review* 29(3), 616–629.
- Dalton, B., E. Glennie, and S. J. Ingels (2009). Late high school dropouts: Characteristics, experiences, and changes across cohorts. descriptive analysis report. nces 2009-307. *National Center for Education Statistics*.
- De Corte, E. (2003). Transfer as the productive use of acquired knowledge, skills, and motivations. *Current directions in psychological science* 12(4), 142–146.
- De Graaf-Zijl, M., G. J. van den Berg, and A. Heyma (2011). Stepping stones for the unemployed: the effect of temporary jobs on the duration until (regular) work. *Journal of Population Economics* 24(1), 107–139.
- De Grip, A. (2009). *Economic perspectives of workplace learning*, Chapter The learning potential of the workplace, pp. 15–29. Sense Publishers.
- De Grip, A., H. Bosma, D. Willems, and M. Van Boxtel (2008). Job-worker mismatch and cognitive decline. *Oxford Economic Papers* 60(2), 237–253.
- De Grip, A. and J. Sauermann (2013). The effect of training on productivity: The transfer of on-the-job training from the perspective of economics. *Educational Research Review* 8, 28–36.
- De Grip, A., J. Sauermann, and I. Sieben (2016). The role of peers in estimating tenure-performance profiles: evidence from personnel data. *Journal of Economic Behavior & Organization* 126, 39–54.
- De Grip, A. and J. Van Loo (2002). The economics of skills obsolescence: a review. In *The economics of skills obsolescence*, pp. 1–26. Emerald Group Publishing Limited.
- De Oliveira, A., R. T. Croson, and C. C. Eckel (2012). Are preferences stable across domains? an experimental investigation of social preferences in the field. *Southern Economic Journal* 79(1), 15–45.
- De Witte, K., S. Cabus, G. Thyssen, W. Groot, and H. M. van den Brink (2013). A critical review of the literature on school dropout. *Educational Research Review* 10, 13–28.
- Dearden, L., H. Reed, and J. Van Reenen (2006). The impact of training on productivity and wages: Evidence from british panel data. *Oxford bulletin of economics and statistics* 68(4), 397–421.
- Destré, G., L. Lévy-Garboua, and M. Sollogoub (2008). Learning from experience or learning from others?: inferring informal training from a human capital earnings function with matched employer–employee data. *The Journal of Socio-Economics* 37(3), 919–938.
- Di Gropello, E. (2006). *Meeting the challenges of secondary education in Latin America and East Asia: Improving efficiency and resource mobilization*. The World Bank.

- Di Pietro, G. and P. Urwin (2006). Education and skills mismatch in the italian graduate labour market. *Applied Economics* 38(1), 79–93.
- Dolton, P. J. and M. A. Silles (2008). The effects of over-education on earnings in the graduate labour market. *Economics of Education Review* 27(2), 125–139.
- Eide, E. R. and M. H. Showalter (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review* 20(6), 563–576.
- European Commission (2007). Towards common principles of flexicurity: more and better jobs through flexibility and security. Technical report, European Commission.
- European Commission (2010a). *Employment in Europe*. Publications Office of the European Union.
- European Commission (2010b). Europe 2020: a strategy for smart, sustainable and inclusive growth. Technical report, European Commission.
- Faccini, R. (2014). Reassessing labour market reforms: temporary contracts as a screening device. *The Economic Journal* 124(575), 167–200.
- Fernandez, R. (2002). The economics of skills obsolescence: Theoretical innovations and empirical applications. In *The economics of skills obsolescence*, pp. 175–193. Emerald Group Publishing Limited.
- Fredriksson, P., B. Öckert, and H. Oosterbeek (2012). Long-term effects of class size. *The Quarterly Journal of Economics* 128(1), 249–285.
- Fruehwirth, J. C., S. Navarro, and Y. Takahashi (2016). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics* 34(4), 979–1021.
- Gagliarducci, S. (2005). The dynamics of repeated temporary jobs. *Labour Economics* 12(4), 429–448.
- García-Pérez, J. I., M. Hidalgo-Hidalgo, and J. A. Robles-Zurita (2014). Does grade retention affect students’ achievement? some evidence from Spain. *Applied Economics* 46(12), 1373–1392.
- Gebel, M. (2013). Is a temporary job better than unemployment? a cross-country comparison based on British, German, and Swiss panel data. *Schmollers Jahrbuch* 133(2), 143–155.
- Gerritsen, S., E. Plug, and D. Webbink (2017). Teacher quality and student achievement: Evidence from a sample of Dutch twins. *Journal of Applied Econometrics* 32(3), 643–660.
- Givord, P. and L. Wilner (2015). When does the stepping-stone work? fixed-term contracts versus temporary agency work in changing economic conditions. *Journal of Applied Econometrics* 30(5), 787–805.

- Görlitz, K. (2011). Continuous training and wages: An empirical analysis using a comparison-group approach. *Economics of education review* 30(4), 691–701.
- Green, F. (1993). The determinants of training of male and female employees in Britain. *Oxford Bulletin of Economics and Statistics* 55(1), 103–122.
- Green, F., D. Ashton, and A. Felstead (2001). Estimating the determinants of supply of computing, problem-solving, communication, social, and teamworking skills. *Oxford Economic Papers* 53(3), 406–433.
- Green, F. and S. McIntosh (2007). Is there a genuine under-utilization of skills amongst the over-qualified? *Applied Economics* 39(4), 427–439.
- Green, F. and Y. Zhu (2010). Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Oxford Economic Papers* 62(4), 740–763.
- Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters* 107(2), 291–296.
- Greene, W. H. (2012). *Econometric analysis, seventh edition*. Pearson Education.
- Groot, W. (1996). The incidence of, and returns to overeducation in the UK. *Applied Economics* 28(10), 1345–1350.
- Guiso, L., P. Sapienza, and L. Zingales (2013). Time varying risk aversion. Technical Report NBER Working paper series No. 19284, National Bureau of Economic Research.
- Hardardottir, H. (2017). Long term stability of time preferences and the role of the macroeconomic situation. *Journal of Economic Psychology* 60, 21–36.
- Hartog, J. (2000). Over-education and earnings: where are we, where should we go? *Economics of education review* 19(2), 131–147.
- Havnes, T. and M. Mogstad (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics* 95(11), 1455–1465.
- Heckman, J. and P. Carneiro (2003). Human capital policy. Technical report, National Bureau of Economic Research.
- Heckman, J., L. Lochner, and R. Cossa (2002). Learning-by-doing vs. on-the-job training: using variation induced by the eitc to distinguish between models of skill formation. Technical Report NBER Working paper series No. 9083, National Bureau of Economic Research.
- Heckman, J. and S. Navarro-Lozano (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics* 86(1), 30–57.
- Heckman, J. J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy* 84(4), 9–44.

- Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4), 931–959.
- Heckman, J. J., L. Lochner, and C. Taber (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents u. *REVIEW OF ECONOMIC DYNAMICS* 1, 158.
- Heckman, J. J., J. L. Tobias, and E. Vytlacil (2001). Four parameters of interest in the evaluation of social programs. *Southern Economic Journal* 68(2), 211–223.
- Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *The American Economic Review* 91(2), 107–111.
- Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73(3), 669–738.
- Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 96(8), 4730–4734.
- Hersch, J. (1991). Education match and job match. *The Review of Economics and Statistics* 73(1), 140–144.
- Hill, A. J. (2014). The costs of failure: Negative externalities in high school course repetition. *Economics of Education Review* 43, 91–105.
- Holmlund, B. and D. Storrie (2002). Temporary work in turbulent times: the swedish experience. *The Economic Journal* 112(480).
- Ikedo, M. and E. García (2014). Grade repetition. *OECD Journal: Economic Studies* 2013(1), 269–315.
- Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47(1), 5–86.
- Jacob, B. A. and L. Lefgren (2004). Remedial education and student achievement: A Regression-Discontinuity analysis. *The Review of Economics and Statistics* 86(1), 226–244.
- Jacob, B. A. and L. Lefgren (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics* 1(3), 33–58.
- Jahn, E. J. and D. Pozzoli (2013). The pay gap of temporary agency workers—does the temp sector experience pay off? *Labour Economics* 24, 48–57.
- Jahn, E. J. and M. Rosholm (2014). Looking beyond the bridge: the effect of temporary agency employment on labor market outcomes. *European Economic Review* 65, 108–125.

- Janosz, M., M. LeBlanc, B. Boulerice, and R. E. Tremblay (1997). Disentangling the weight of school dropout predictors: A test on two longitudinal samples. *Journal of youth and adolescence* 26(6), 733–762.
- Jimerson, S., E. Carlson, M. Rotert, B. Egeland, and L. A. Sroufe (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology* 35(1), 3–25.
- Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology* 37(3), 243–272.
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School psychology review* 30(3), 420–437.
- Jimerson, S. R., G. E. Anderson, and A. D. Whipple (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools* 39(4), 441–457.
- Kahn, L. M. (2010). Employment protection reforms, employment and the incidence of temporary jobs in europe: 1996–2001. *Labour Economics* 17(1), 1–15.
- Karaca-Mandic, P., E. C. Norton, and B. Dowd (2012). Interaction terms in nonlinear models. *Health services research* 47(1pt1), 255–274.
- Kattan, R. B. and M. Székely (2015). *Analyzing the dynamics of school dropout in upper secondary education in Latin America: a cohort approach*. The World Bank.
- Killingsworth, M. R. (1982). “learning by doing” and “investment in training”: a synthesis of two “rival” models of the life cycle. *The Review of Economic Studies* 49(2), 263–271.
- Koopmans, H., A. J. Doornbos, and I. M. v. Eekelen (2006). Learning in interactive work situations: it takes two to tango; why not invite both partners to dance? *Human Resource Development Quarterly* 17(2), 135–158.
- Koppensteiner, M. F. (2014). Automatic grade promotion and student performance: Evidence from Brazil. *Journal of Development Economics* 107, 277–290.
- Lamb, S. and E. Markussen (2011). School dropout and completion: An international perspective. In *School dropout and completion*, pp. 1–18. Springer.
- Lamote, C., S. Speybroeck, W. Van Den Noortgate, and J. Van Damme (2013). Different pathways towards dropout: the role of engagement in early school leaving. *Oxford Review of Education* 39(6), 739–760.
- Leuven, E. and H. Oosterbeek (2008). An alternative approach to estimate the wage returns to private-sector training. *Journal of applied econometrics* 23(4), 423–434.

- Levitt, S., J. List, C. Syverson, et al. (2012). Toward an understanding of learning by doing: evidence from an automobile plant. Technical report, The Field Experiments Website.
- Linsley, I. et al. (2005). Causes of overeducation in the Australian labour market. *Australian Journal of Labour Economics* 8(2), 121–144.
- Loewenstein, M. A. and J. R. Spletzer (1998). Dividing the costs and returns to general training. *Journal of Labor Economics* 16(1), 142–171.
- Loh, E. S. (1994). Employment probation as a sorting mechanism. *Industrial and Labor Relations Review* 47(3), 471–486.
- Long, J. (1997). Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences* 7.
- Long, J. S. and J. Freese (2006). *Regression models for categorical dependent variables using Stata*. Stata press.
- Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.
- Mahjoub, M.-B. (2017). The treatment effect of grade repetitions. *Education Economics* 25(4), 418–432.
- Malmendier, U. and S. Nagel (2011). Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics* 126(1), 373–416.
- Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *The Quarterly Journal of Economics* 131(1), 53–87.
- Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics* 94(2), 596–606.
- Marsick, V., K. Watkins, M. Callahan, and M. Volpe (2015). *Informal and incidental learning in the workplace*, Chapter Handbook of research on adult learning and development, pp. 570–599. Routledge.
- Martínez, G. and B. Herrera (2002). Finalidades y alcances del decreto 230 del 11 de febrero de 2002. *Ministerio de Educación Nacional. Bogotá*, 89–90.
- Maurer, T. J. and E. M. Weiss (2010). Continuous learning skill demands: associations with managerial job content, age, and experience. *Journal of Business and Psychology* 25(1), 1–13.
- Mavromaras, K. and S. McGuinness (2012). Overskilling dynamics and education pathways. *Economics of Education Review* 31(5), 619–628.
- Mavromaras, K., S. McGuinness, N. O’leary, P. Sloane, and Y. K. Fok (2010). The problem of overskilling in Australia and Britain. *The Manchester School* 78(3), 219–241.

- McCoy, A. R. and A. J. Reynolds (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology* 37(3), 273–298.
- McGuinness, S. (2006). Overeducation in the labour market. *Journal of economic surveys* 20(3), 387–418.
- McGuinness, S. and D. Byrne (2014). Examining the relationships between labour market mismatches, earnings and job satisfaction among immigrant graduates in europe. Technical Report 8440, IZA.
- McGuinness, S. and P. J. Sloane (2011). Labour market mismatch among uk graduates: An analysis using reflex data. *Economics of Education Review* 30(1), 130–145.
- McGuinness, S. and M. Wooden (2009). Overskilling, job insecurity, and career mobility. *Industrial relations: a journal of economy and society* 48(2), 265–286.
- McLanahan, S. and G. Sandefur (1994). *Growing Up with a Single Parent. What Hurts, What Helps*. ERIC.
- Meier, S. and C. D. Sprenger (2015). Temporal stability of time preferences. *Review of Economics and Statistics* 97(2), 273–286.
- Messinis, G. and N. Olekalns (2007). Skill mismatch and training in australia: Some implications for policy. *Australian Economic Review* 40(3), 300–306.
- Messinis, G. and N. Olekalns (2008). Returns to training and skill mismatch: evidence from australia. Technical Report 40, Victoria University.
- Mincer, J. (1962). On-the-job training: Costs, returns, and some implications. *Journal of political Economy* 70(5, Part 2), 50–79.
- Mincer, J. (1974). *Schooling, experience, and earnings*. Columbia University Press.
- Mincer, J. (1997). The production of human capital and the life cycle of earnings: variations on a theme. *Journal of Labor Economics* 15(1), 26–47.
- Mincer, J. and H. Ofek (1982). Interrupted work careers: Depreciation and restoration of human capital. *Journal of human resources* 17(1), 3–24.
- Miranda, A. and S. Rabe-Hesketh (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, count, and ordinal variables. *Stata Journal* 6(3), 285–308.
- Nagypál, É. (2007). Learning by doing vs. learning about match quality: can we tell them apart? *The Review of Economic Studies* 74(2), 537–566.
- Norton, E. C., H. Wang, C. Ai, et al. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal* 4(2), 154–167.

- O'Connell, P. J. and D. Byrne (2010). The determinants and effects of training at work: bringing the workplace back in. *European Sociological Review* 28(3), 283–300.
- OECD (2004). *OECD Employment Outlook 2004*. OECD Publishing.
- OECD (2010). *Recognition of non-formal and informal learning: country practices*. OECD Publishing.
- OECD (2011). *Upper Secondary Completion: an OECD Overview*. OECD Publishing.
- OECD (2012). *Education Indicators in Focus*. OECD Publishing.
- OECD (2013). *Skills for life? Key findings from the survey of adult skills*. OECD Publishing.
- OECD (2014). *OECD Employment Outlook 2014*. OECD Publishing.
- OECD (2016). *Education in Colombia Highlights 2016*. OECD Publishing.
- O'Leary, N., P. Sloane, S. McGuinness, P. O'Connor, and K. Mavromaras (2009). A taxonomy of skill mismatch. Technical report, European Centre for the Development of Vocational Training (Cedefop).
- Pharris-Ciurej, N., C. Hirschman, and J. Willhoft (2012). The 9th grade shock and the high school dropout crisis. *Social science research* 41(3), 709–730.
- Plank, S., D. S. and E. A. (2005). Dropping out of high school and the place of career and technical education: a survival analysis of surviving high school. Technical report, National Research Center for Career and Technical Education: Atlanta, GA.
- Psacharopoulos, G. (2007). *The costs of school failure: A feasibility study*. EENEE Brussels.
- Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association* 67(338), 306–310.
- Quiñones, M. A., J. K. Ford, and M. S. Teachout (1995). The relationship between work experience and job performance: a conceptual and meta-analytic review. *Personnel Psychology* 48(4), 887–910.
- Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128(2), 301–323.
- Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica* 73(2), 417–458.
- Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal* 31(4), 729–759.

- Rosen, S. (1972). Learning and experience in the labor market. *Journal of Human Resources* 7, 326–342.
- Rosen, S. (1975). Measuring the obsolescence of knowledge. In *Education, income, and human behavior*, pp. 199–232. NBER.
- Rumberger, R. (2011). *Dropping Out: Why Students Drop Out of High School and What Can Be Done About It*. Cambridge, MA Harvard University Press.
- Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American educational Research journal* 32(3), 583–625.
- Rumberger, R. W. (2004). *Dropouts in America: Confronting the Graduation Rate Crisis*, Chapter Why students drop out school, pp. 131–155. Harvard Education Press, Cambridge.
- Rumberger, R. W. and S. P. Lamb (2003). The early employment and further education experiences of high school dropouts: A comparative study of the united states and australia. *Economics of Education Review* 22(4), 353–366.
- Rumberger, R. W. and K. A. Larson (1998). Student mobility and the increased risk of high school dropout. *American journal of Education* 107(1), 1–35.
- Sala, H., J. I. Silva, and M. Toledo (2012). Flexibility at the margin and labor market volatility in oecd countries. *The Scandinavian Journal of Economics* 114(3), 991–1017.
- Sepúlveda, F. (2009). Training and productivity: evidence for us manufacturing industries. *Oxford Economic Papers* 62(3), 504–528.
- Sicherman, N. and O. Galor (1990). A theory of career mobility. *Journal of Political Economy* 98(1), 169–192.
- Silberglitt, B., J. J. Appleton, M. K. Burns, and S. R. Jimerson (2006). Examining the effects of grade retention on student reading performance: A longitudinal study. *Journal of School Psychology* 44(4), 255–270.
- Sloane, P. J., H. Battu, and P. T. Seaman (1996). Overeducation and the formal education/experience and training trade-off. *Applied Economics Letters* 3(8), 511–515.
- Sloane, P. J., H. Battu, and P. T. Seaman (1999). Overeducation, undereducation and the british labour market. *Applied Economics* 31(11), 1437–1453.
- Stearns, E. and E. J. Glennie (2006). When and why dropouts leave high school. *Youth & Society* 38(1), 29–57.
- Stearns, E., S. Moller, J. Blau, and S. Potochnick (2007). Staying back and dropping out: The relationship between grade retention and school dropout. *Sociology of Education* 80(3), 210–240.

- Steijn, B., A. Need, and M. Gesthuizen (2006). Well begun, half done? long-term effects of labour market entry in the netherlands, 1950-2000. *Work, Employment and Society* 20(3), 453–472.
- Swanson, C. B. (2004). Sketching a portrait of public high school graduation: Who graduates? who doesn't. *Dropouts in America: Confronting the graduation rate crisis*, 13–40.
- Temple, J., A. Reynolds, and S.-R. Ou (2004). Grade retention and school dropout: Another look at the evidence: Grade retention, tracking and grouping. In *Can unlike students learn together? Grade retention, tracking and grouping: Grade retention, tracking and grouping*. Information Age.
- Tesluk, P. E. and R. R. Jacobs (1998). Toward an integrated model of work experience. *Personnel psychology* 51(2), 321–355.
- Van Smoorenburg, M. and R. K. Van der Velden (2000). The training of school-leavers: complementarity or substitution? *Economics of education review* 19(2), 207–217.
- Verhaest, D. and E. Omeij (2009). Objective over-education and worker well-being: A shadow price approach. *Journal of Economic Psychology* 30(3), 469–481.
- Virtanen, M., M. Kivimäki, M. Joensuu, P. Virtanen, M. Elovainio, and J. Vahtera (2005). Temporary employment and health: a review. *International Journal of Epidemiology* 34(3), 610–622.
- Wasmer, E. (1999). Competition for jobs in a growing economy and the emergence of dualism. *The Economic Journal* 109(457), 349–371.
- Weiss, Y. (1986). *The determination of life cycle earnings: a survey*, Volume 1, Chapter Handbook of Labour Economics, pp. 603–640. Elsevier.
- Welch, F. and M. Ureta (2002). The obsolescence of skill. In *The Economics of Skills Obsolescence*, pp. 51–81. Emerald Group Publishing Limited.
- Winship, C. and R. D. Mare (1984). Regression models with ordinal variables. *American sociological review* 49(4), 512–525.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- World Economic Forum (2014). *Matching Skills and Labor Market Needs: Building Social Partnerships for Better Skills and Better Jobs*, Volume 1. Global agenda council on employment, World Economic Forum.

About the author

Maria Ferreira Sequeda was born 11 April 1984 in San Gil, Colombia. She obtained her bachelor degree in Economics, *cum laude*, from the Industrial University of Santander (2006) and a master degree in Development Management from the Andes University (2010). She held a teaching and research position at the Economics Department of the Industrial University of Santander from 2007 to 2017. Maria obtained a second master degree in Development Studies, *with distinction*, from Erasmus University Rotterdam in 2012. She joined the Research Centre for Education and the Labour Market (ROA) at Maastricht University in 2013 as a PhD candidate. Since January 2018, Maria works as a Senior Economist and Research Professional at the global Research Department of ING Bank.

ROA Dissertation Series

1. Lex Borghans (1993). *Educational Choice and Labour Market Information*. Maastricht, Research Centre for Education and the Labour Market.
2. Frank Cörvers (1999). *The Impact of Human Capital on International Competitiveness and Trade Performance of Manufacturing Sectors*. Maastricht, Research Centre for Education and the Labour Market.
3. Ben Kriechel (2003). *Heterogeneity Among Displaced Workers*. Maastricht, Research Centre for Education and the Labour Market.
4. Arnaud Dupuy (2004). *Assignment and Substitution in the Labour Market*. Maastricht, Research Centre for Education and the Labour Market.
5. Wendy Smits (2005). *The Quality of Apprenticeship Training, Conflicting Interests of Firms and Apprentices*. Maastricht, Research Centre for Education and the Labour Market.
6. Judith Semeijn (2005). *Academic Competences and Labour Market Entry: Studies Among Dutch Graduates*. Maastricht, Research Centre for Education and the Labour Market.
7. Jasper van Loo (2005). *Training, Labor Market Outcomes and Self-Management*. Maastricht, Research Centre for Education and the Labour Market.
8. Christoph Meng (2005). *Discipline-Specific or Academic? Acquisition, Role and Value of Higher Education Competencies*. Maastricht, Research Centre for Education and the Labour Market.
9. Andreas Ammermüller (2007). *Institutional Effects in the Production of Education: Evidence from European Schooling Systems*. Maastricht, Research Centre for Education and the Labour Market.
10. Bart Golsteyn (2007). *The Ability to Invest in Human Capital*. Maastricht, Research Centre for Education and the Labour Market.
11. Raymond Montizaan (2010). *Pension Rights, Human Capital Development and Well-Being*. Maastricht, Research Centre for Education and the Labour Market.
12. Annemarie Nelen (2012). *Part-Time Employment and Human Capital Development*. Maastricht, Research Centre for Education and the Labour Market.

13. Jan Sauermann (2013). *Human Capital, Incentives, and Performance Outcomes*. Maastricht, Research Centre for Education and the Labour Market.
14. Harald Ulrich Pfeifer (2013). *Empirical Investigations of Costs and Benefits of Vocational Education and Training*. Maastricht, Research Centre for Education and the Labour Market.
15. Charlotte Büchner (2013) *Social Background, Educational Attainment and Labor Market Integration: An Exploration of Underlying Processes and Dynamics*. Maastricht, Research Centre for Education and the Labour Market.
16. Martin Humburg (2014). *Skills and the Employability of University Graduates*. Maastricht, Research Centre for Education and the Labour Market.
17. Jan Feld (2014). *Making the Invisible Visible, Essays on Overconfidence, Discrimination and Peer Effects*. Maastricht, Research Centre for Education and the Labour Market.
18. Olga Skriabikova (2014). *Preferences, Institutions, and Economic Outcomes: an Empirical Investigation*. Maastricht, Research Centre for Education and the Labour Market.
19. Gabriele Marconi (2015). *Higher Education in the National and Global Economy*. Maastricht, Research Centre for Education and the Labour Market.
20. Nicolas Salamanca Acosta (2015). *Economic Preferences and Financial Risk-Taking*. Maastricht, Research Centre for Education and the Labour Market.
21. Ahmed Elsayed Mohamed (2015). *Essays on Working Hours*. Maastricht, Research Centre for Education and the Labour Market.
22. Roxanne Amanda Korthals (2015). *Tracking Students in Secondary Education, Consequences for Student Performance and Inequality*. Maastricht, Research Centre for Education and the Labour Market.
23. Maria Zumbuehl (2015). *Economic Preferences and Attitudes: Origins, Behavioral Impact, Stability and Measurement*. Maastricht, Research Centre for Education and the Labour Market.
24. Anika Jansen (2016). *Firms' Incentives to Provide Apprenticeships – Studies on Expected Short- and Long-Term Benefits*. Maastricht, Research Centre for Education and the Labour Market.
25. Jos Maarten Arnold Frank Sanders (2016). *Sustaining the Employability of the Low-Skilled Worker: Development, Mobility and Work Redesign*. Maastricht, Research Centre for Education and the Labour Market.

26. Marion Collewet (2017). *Working Hours: Preferences, Well-Being and Productivity*. Maastricht, Research Centre for Education and the Labour Market.
27. Tom Stolp (2018). *Sorting in the Labor Market: The Role of Risk Preference and Stress*. Maastricht, Research Centre for Education and the Labour Market.
28. Frauke Meyer (2019). *Individual Motives for (Re-)Distribution*. Maastricht, Research Centre for Education and the Labour Market.
29. Maria Ferreira Sequeda (2019). *Human Capital Development at School and Work*. Maastricht, Research Centre for Education and the Labour Market.