UNIVERSITAT DE BARCELONA

Analysis of the process of acquisition of skills and competencies in Spain

Juan Manuel Guio Jaimes

2017

PhD in Economics | Juan Manuel Guio Jaimes

UNIVERSITAT DE
BARCELONA

PhD in Economics

Analysis of the process of acquisition of skills and competencies in Spain

Juan Manuel Guio Jaimes

PhD in Economics

Thesis title:

Analysis of the process of acquisition of skills and competencies in Spain

PhD student:
Juan Manuel Guio Jaimes

Advisors:
Álvaro Choi de Mendizábal
Josep-Oriol Escardíbul Ferra

Date:
June 2017

UNIVERSITAT DE BARCELONA

Acknowledgements

There are various people without whom this thesis may never have been written. To all of them, I would like to express my gratitude.

First, I would like to thank my advisor, Álvaro Choi. I am deeply indebted to Álvaro for the fundamental role he has played throughout the writing of my doctoral thesis. He has provided me with all the guidance, assistance and expertise that I have needed during these years. When I felt ready to venture out on my own into new research areas, Álvaro gave me the freedom to follow my own academic interests, but at the same time continued to offer invaluable feedback, advice and encouragement. In addition to this academic collaboration, I greatly value the personal rapport we have forged over these years. I cannot possibly imagine a better advisor.

I would also like to thank Oriol Escardíbul Ferrá. His knowledge and advice have helped me to stay on track and to work at a smooth and regular pace. His careful revisions of my essays have honed my research and organizational skills and made my writing a little less amateurish. My gratitude is extended as well to all the professors in the Faculty of Economics for their advice and insights kindly offered over these years and for the substantial influence that their courses have had on my research. Finally, I gratefully acknowledge the time dedicated to this thesis and the valuable feedback received from all the members of my PhD committee.

To my mother, Jenny, for her unconditional love throughout the most difficult moments of my life. To my wife, Olga, for her constant support and patience and for her numerous sacrifices throughout this long journey that has not always been easy. To my loving daughter, Montserrat, for her care and kindness that helped me overcome all the obstacles in my pursuit of a PhD degree. To Matias for his faithful company during long sleepless hours. Thank you.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 General topic and background

Education is one of the principal mechanisms to intervene in the process of human capital accumulation and, as such, it has a significant impact on economic growth and on the improvement of life conditions. The importance of the role that formal academic instruction plays in the development process of a country and of its population has generated widespread concern to identify policies that can guarantee the provision of high-quality education. Indeed, recent evidence seems to show an increasing demand for education and competencies (Rakowska and Cichorzewska, 2016) [120].

The provision of education is challenged by limited resources so that different educational technologies result in different outcomes. This means that the diverse results produced by educational policies around the world are, to some extent, attributable to differences in educational inputs (Hanushek, 1997) [67].

However, differences in academic outcomes cannot be explained solely by the characteristics of educational systems (Coleman et al., 1966) [44]. In fact, more often than not, such differences have been defined long before the students enter the school (Lee and Burkham, 2002) [81], as a large number of personal, household and school level variables have an important bearing on the performance and permanency of students in a school system. As such, it is a pressing concern for policymakers to determine which of these factors have the most significant impact on academic results.

Indeed, the situation is yet more complex as there are many other possible sources of variation in academic outcomes that extend beyond the schools, families and the students' personal characteristics. Factors such as labour market dynamics (Aparicio, 2010) [9], economic cycles (Malley and Muscatelli, 1999) [86] and the expansion of the productive

sectors of the economy, among others, can have an impact on educational decisions. However, these factors have normally only been explored at a descriptive level, and rarely from an analytic perspective. Yet, it is of great interest to analyse the effects of household, personal and school characteristics, in addition to the influence of labour market characteristics, on educational decisions.

Students naturally progress from the education system into the labour market, having acquired a certain degree of formal education and having developed a certain level of competencies, all dependent on their schooling decisions, their abilities and their experience. Here, a number of factors can be considered relevant in determining their expected remuneration in terms of wages, above all the closeness of their match with the occupational requirements needed to perform the job, and their individual occupational endowments.

In this sense, while it is well established that, on average, more highly educated workers tend to have higher salaries (Patrinos, 2016) [112], there may be situations in which this relationship does not necessarily hold, as salary levels are dependent on the interaction of a set of labour demand and supply factors. Technological innovations, changes in labour market structures, and changes in the distribution of education and skills can all affect the demand and supply of occupational characteristics and, as a result, the structure of wages across time. Wage inequality has been a major concern in recent decades, often associated with changes in demand that favour more educated and skilled workers, or with changes in the socio-demographic composition of countries, with foreign workers competing with local workers or with an increase in female participation in the labour market (Bloom et al., 2009) [23]. Individuals are remunerated according to their level of education and competencies, or by the level of education and competencies required by a particular job. Individuals with excess education or competencies (so-called mismatch) can often find themselves penalized in comparison to those that have the required level of education to perform a given task, while workers with an insufficient level of education or competencies for their occupation usually obtain a premium in their wage return (Flisi et al., 2014) [60].

However, the relationship between the level of education/competencies required to perform a job and the education/competencies possessed by an individual and the way in which the labour market penalizes/rewards the mismatches between the supply and demand sides of these skills are not always clear.

It therefore follows that differences in the remuneration of workers may, in part, be explained by education and skills mismatches. Considering that there is an enduring wage gap between males and females (Angelov et al., 2014) [7], it is clearly of interest to analyse if these persistent differences in remuneration can be justified in terms of discrepancies in the

occupational endowments of men and women and the level of education and skills required to carry out the job.

Given the important impact of an individual's education and skills on their life trajectory, it is clearly worthwhile establishing the determinants of the acquisition of competencies, and the way in which educational decisions are affected by personal, household and school characteristics (Paper 1 – Chapter 2); the way in which they are affected by the economic environment and by labour market dynamics (Paper 2 – Chapter 3); and the effect that these schooling decisions have on workers' expected returns. The latter, and their heterogeneous effects across countries and gender, are analysed in Paper 3 – Chapter 4.

1.2 Literature review

The literature examining the educational production function – the main theoretical framework underpinning this thesis – is broad and already has a long trajectory. Most significantly, in this regard, Hanushek (2003) [68] has proposed an educational production function that distinguishes between personal, household and school characteristics.

An obvious personal characteristic that would appear to have an important impact on academic performance and permanency is the gender of the student. When considering reading competency tests, women traditionally perform better than men, while males usually obtain better results than women on tests of mathematical competence (Calero and Choi, 2010 [27]; Fernandez Enguita, 2010) [58]. Various authors, including Bertrand and Pan (2011) [19] and Camarata and Woodcock (2006) [31], conclude that, in general, females are more mature than males and that the former are more successful in the school system.

As Crawford et al. (2011) [47], García Montalvo (2011) [63], and Robertson (2011) [125] all point out, the month in which a person is born may also have a marked impact on their expected academic outcomes. Students born during the first quarter of the year seem to perform better than those born between October and December. Other personal characteristic that have been examined include grade retention, with a negative relationship being shown with educational results (Holmes, 1989 [71]; Jimmerson, 2001 [74]), and being born outside the country, which is positively associated with being at risk of school failure (Calero and Waisgrais, 2009 [29]; McCarthy, 1998 [90]).

Turning to household characteristics, a family's socioeconomic background seems to be very closely related to an individual's academic performance and to their school permanency. As the Coleman Report (1966) [44], Hanushek (1997) [67] and Haveman and Wolfe (1995) [69] conclude, family background is one of the main factors influencing a students'

academic achievement. The higher the educational level (Ferguson et al., 1996) [57] and professional category (Cohen, 1987) [43] of the parents, the lower the student's risk of school failure. Likewise, the availability of cultural resources (Berger et al., 2005) [18] and the possession of a significant number of books in the household – a proxy for the socio-cultural environment (Aikens and Barbarin, 2008) [2] – positively affect school permanency. The family's country of origin is also relevant, as Brunello and Rocco (2013) [24] show that first generation immigrants tend to have a lower performance.

School attributes can enhance or diminish academic performance and modulate the effect that personal and household characteristics have on the decision of young adults to further their education. In the Spanish case, school ownership does not seem to have a statistically significant effect. The composition of the school population, that is the proportion of students by gender and by immigrant status, appears to have a marked effect on students' academic outcomes (Calero and Escardíbul, 2007 [28]; Calero and Waisgrais, 2009 [29]; Cordero, 2011 [46]). Other school characteristics, including the student-teacher ratio, present contradictory evidence. For Chingos (2010) [40] and Hanushek (2003) [68], this ratio does not appear to have any impact on academic performance, while for Angrist and Lavy (1999) [8] reducing the class size results in a significant increase in students' test scores. Autonomy as regards budget allocation and school content does not have a particular impact on students' expected outcomes (Calero and Waisgrais, 2009) [29].

Labour market dynamics, especially during economic downturns, are recognized as an important determinant of individual educational decisions (Black et al., 2005) [21]. If unemployment is widespread, two outcomes are possible: on the one hand, if students perceive a deterioration in the financial conditions of the household – the so-called “family" effect of unemployment, they may opt to abandon their schooling in order to seek employment and contribute to the household income (Carneiro and Heckman, 2002 [33]; Entwisle et al., 2005 [53] ; Montmarquette et al., 2007 [98]). As such, the “family" effect of unemployment has a negative impact on the individual demand for education.

On the other hand, if finding or maintaining a job becomes difficult for the students' parents, their relatives, the parents of their schoolfellows and other acquaintances, the students may perceive the adverse economic conditions and become aware of the general lack of job opportunities – the so-called “local labour market" effect of unemployment. This effect encourages students to remain in the school system and, as such, has a positive effect on the demand for education (Bickel, 1989 [20]; McNeal, 1997 [94]; Rees and Mocan, 1997 [122] and Rivkin, 1995 [124]).

However, other authors, including Card and Lemieux (2001) [32], Manski and Wise (1983) [87], Schady (2004) [132] and Warren and Lee (2003) [141], do not find any significant evidence of a distinctive effect of labour market conditions on young adults' schooling decisions. In this sense, disentangling the family and local labour market effects and their repercussions on the individual demand for education seems to be an important step for understanding educational decisions.

Sooner or later, individuals leave the educational system and, most of them, become part of the labour force. Finding a job may be seen as the first step in this process, but the quality of employment is also relevant, with the wage received being a key indicator. While some of the factors that determine these outcomes lie beyond the control of the individual – including the education of their parents, the labour market environment and even their gender – the individual has, to some extent, a certain amount of freedom to take individual decisions concerning their level of education. Of course, this level of freedom may be severely conditioned by individual or family characteristics. Indeed, inequalities at early stages in life may lead to the reproduction of these inequalities in adult life.

All in all, educational attainment and skills acquisition are important determinants of wages (Quinn and Rubb, 2006) [117]. Consequently, differences in the level of education or the level of skills, or differences in the effect of education or skills mismatches may be important drivers of wage gaps. These may be particularly significant if we take into consideration the persistent unexplained wage inequalities between genders (Robst, 2007 [126]; Salinas-Jiménez et al. 2013 [129]).

At least four theories have been forwarded in explanation of occupational mismatches. First, human capital theory found a direct link between investment in the education of individuals, their productivity and the expected return on this investment as a wage – the higher the investment, the higher the return (Becker, 1964) [16]. Second, the signalling/screening theory emphasized that although individual education decisions might be seen as a beneficial investment for workers, the investment may have no effect on expected on-the-job productivity (Spence, 1973) [135]. Third, the job competition theory posited that employers, in an asymmetric information framework, take the educational level of individuals as an indicator of the cost of training to perform a specific job. In this framework, workers may accept positions for which they are overeducated while looking for better job prospects (Thurow, 1975) [139]. Finally, labour market allocation models affirm that the quality of matching between heterogeneous individuals to heterogeneous work positions determines the wage return to the education investment (Sattinger, 1993) [131].

Educational mismatches describe the market misallocations between the workers' level of education and the education required to perform the job. Most of the literature suggests that overeducated workers suffer a penalization on their wage return compared to that of workers holding the same position that are well matched, while undereducated workers usually obtain a premium (Chevalier, 2003 [39]; Dolton and Vignoles, 2000 [50]; Groot, 1996 [66]). Education and skill mismatches are closely linked but represent distinct phenomena. Thus, skill mismatches measure a more dynamic feature that includes the competencies and abilities possessed by a worker and those required to function correctly in the workplace. In this sense, skill mismatches have not been as widely explored as educational mismatches – due essentially to data scarcity, but most of the literature contends that skill mismatches have a negative effect on earnings, albeit of a lower magnitude than that attributable to educational mismatch (Allen and van der Velden, 2001 [4]; Green et al., 2007 [65]; Robst, 2007 [126]). While most of the literature has analysed the effect of educational mismatch at the country level (see Nordin et al., 2010 [99], for example) and a number of papers have focused on skills mismatches (Liu et al., 2016 [83], and Slonimczyk, 2013 [134], being good examples), a recent strand in the literature has performed comparative analyses of skills mismatches. For example, Santos and Sequeira (2014) [130] found a positive effect of underskilling and a negative effect of overskilling on wages, while Budría and Moro-Egido (2014) [26] claim that the wage differences of matched and mismatched workers are not consistent across the income distribution.

Even rarer are studies that examine the gender wage gap of occupationally mismatched workers. Salinas-Jimenez et al. (2013) [129] found that the wage gap is lower for women with less education. Robst (2007) [126] reported that men are more likely to be mismatched than women, and Addabbo and Favaro (2011) [1] concluded that overeducated females are the segment of the population for whom the differences in wage returns by gender are most extreme. To the best of our knowledge, no studies to date have examined the gender wage gap of occupationally mismatched workers across countries.

1.3 Research objectives

The objectives of this thesis are:

1. To describe and provide evidence of the effect of personal, household and school characteristics on the risk of school failure in Spain during the first decade of the new millennium (Chapter 2).

2. To analyse the links between labour market conditions, academic performance and students' decisions to remain in the educational system (Chapter 3).

3. To provide evidence of the link between occupational mismatch and the gender wage gap and to explore the factors underpinning these wage differentials across countries (Chapter 4).

1.4 State of the research

Objective one: School failure in Spain appears to be attributable to a number of structural characteristics, as the phenomenon has been present in the country's educational system for more than 30 years, with dropout rates fluctuating around 30% over the last two decades. Failure to accomplish the objective set by the 2010 Lisbon Strategy and difficulties in achieving the Europe 2020 Strategy of reducing the early school leaver rate to less than 10% are clear indicators of the problems Spain has encountered in addressing the underlying causes of school failure.

Studies such as Calero, Choi and Waisgrais (2010) [27] and Fernández, Mena and Riviere (2010) [58] have analysed the determinants of the risk of school failure in Spain. However, the evolution taken by these variables during the first decade of the new millennium has yet to be studied. From a policy point of view, it is crucial not only to identify these determinants and their magnitude, but also to track shifts and trends in them.

To achieve this objective, a number of research questions have to be addressed, based on the extensive literature discussing the determinants of academic performance. For example: How significant is the impact of the student's socio-economic background and the educational achievements of their parents? How important is the student's gender in their probability of being at risk of school failure? What is the impact of immigration on the probability of school failure? How relevant a factor is school autonomy? Do students enrolled in public schools face a higher risk of school failure?

Likewise, it is relevant to identify trends in the impact of personal, household and school characteristics on the probability of a student obtaining a score below level-2 on the test in reading competence organized by the Programme for International Student Assessment (PISA). In other words, by addressing this objective we should provide answers to the following questions: What determinants were significant throughout the period? Can any specific trends be identified? And, naturally, what reasons might explain the trends observed?

Objective two: Labour market conditions and job opportunities seem to be closely related to the educational decisions taken by individuals. Studies, such as Petrongolo and San Segundo (2002) [115], focus on the relationship between dropout decisions and labour market shocks. The relationship between education and the labour market is, however, complex. For instance, the labour market generally rewards more positively individuals with more schooling, as an indicator of their greater skill levels; yet, the monetary returns to education tend to fall when a large majority of the population reaches a certain education level (Knight and Sabot, 1987) [76]. At the same time, during economic crises, unemployment rates grow and the monetary returns to education might decrease, significantly affecting the population with the lowest levels of formal instruction. Indeed, rising unemployment rates may have a double effect on educational decision-making: On the one hand, they lower the opportunity cost of studying; on the other, they may force students to find a job so as to complement their household income.

The effect of the Spanish economic downturn on school dropout decisions has been investigated in various studies, including Peraita and Pastor (2000) [113] and Petrongolo and San Segundo (2002) [115], but most of them do not rise above the descriptive level. As such, it is of particular interest to analyse the impact of the labour market situation on the academic performance of Spanish students at the age of 15.

It is particularly insightful to explore the influence of Spanish labour market dynamics on academic performance. Accordingly, here, this study places special emphasis on a set of research questions that include: Do labour market conditions affect academic performance at age 15? Do improved labour market conditions increase the risk of school failure?

Additionally, it seems valuable to identify the heterogeneous effects of the labour market situation on students that present different profiles. If the impact that labour market decisions have on academic performance operates through the logic of opportunity costs and the need to boost household income, it is highly likely that this impact will vary according to the socio-economic characteristics of the household. Therefore, we are interested in analysing if the labour market situation affects all students equally.

Traditionally, individuals have tended to concentrate their learning efforts during their formal years of schooling, which usually occurs during a specific period of their lives (Claxton and Lucas, 2009) [42]. However, the need to develop key skills and competencies during each specific moment in life requires that this process be extended beyond the school training of individuals. This process ensures a greater degree of social cohesion (Fryer, 2016) [62] and allows workers to acquire other skills that are highly valued by the labour market.

When individuals make the transition from the school system to the labour market, their wages are determined mainly by their productivity, but also by the amount of education they have received or their possession of certain abilities. Most of the literature concludes that overeducated workers face a penalization in their wage returns compared to those of adequately matched workers, whose education level equals the educational demands of their jobs. Less overwhelming results have shown that in most scenarios workers with an excess of skills for the requirements of the job they perform are often penalized compared to well-matched workers in terms of their possession of the competencies and skills needed to perform the work.

Objective three: Overeducation and overskilling and their influence on wages have received much attention of late. However, the returns to occupational mismatch and their relation with gender wage differentials have not been extensively studied. Few papers to date have explored the effect of education mismatch on gender inequality and even fewer the relation between skill mismatch and gender wage differentials. However, to the best of our knowledge, no studies have attempted the simultaneous analysis of the effects of overeducation and overskilling on gender wage differentials in the framework of a comparative analysis between countries.

It would appear to be highly pertinent to examine whether the incidence of occupational mismatch has a differentiated impact depending on gender. Addressing this objective will enable us to answer a number of research questions, including: Does wage penalization as a result of occupational mismatch have a greater impact on women in every type of mismatch? If occupational mismatch affects females more significantly, does this account for the unexplained wage differentials between males and females?

The effects of the mismatch between skills and labour market requirements on the gender wage gap appear to vary considerably more across Europe countries than do the effects of education mismatches (Santos and Sequeira, 2013) [130]. Therefore, it is of obvious interest to determine whether there are common patterns in these countries that might shed light on the factors responsible for these gender inequalities, and so provide answers to such questions as: Do occupational mismatches correlate strongly with gender wage differentials? Does the empowerment of women and public social expenditure have an effect on gender discrimination? Can common patterns be identified by country?

1.5 Outline of the current situation

Most of the analyses performed in this thesis focus on the Spanish case, which presents various characteristics that make it especially suitable for the analysis of early school dropout. Indeed, Spain is the country with the highest school dropout rates in the EU. According to Eurostat data, in 2015, approximately 20.0% of Spaniards aged 18 to 24 drop out early from the education system. This means that school dropout rates in Spain are almost double the EU average rate (11.0%). In 2006, almost 30.3% of Spanish students abandoned the education system early, and by the end of 2014 this number stood at around 21.9%. Early school dropout rates are higher for males (25.6 vs 18.1%, for females). And while early leaver rates are slowly receding, school failure is still a major concern in Spanish society today.

Between 1996 and 2007, Spain experienced a period of intense economic growth and development, with a marked boost in labour opportunities for young adults. Jobs for low-skilled workers were also generated, above all in the construction and service sectors (Arranz and García-Serrano, 2012) [10]. From 2000 to 2007, the general rate of unemployment fell from 10.6 to 8.3%. After that however, the country was hit by a severe economic crisis and, by 2008, the adult unemployment rate had reached 11.3%. Thereafter the situation deteriorated rapidly and the general unemployment rate peaked at 24.8% in 2012.

According to the Spanish Statistical Office (INE), the unemployment rate for young adults (below 24 years of age) in the first quarter of 2017 was around 41.7%, more than double the general unemployment rate (17.6%). Since 2000, the school dropout rates appear to have risen during periods of economic expansion and to have receded during times of economic unrest. This seems to point to a negative relationship between unemployment and school dropout rates.

In parallel with this, according to Eurostat the gender wage gap in Spain climbed from 18.1% in 2007 to 18.8% in 2014, suggesting that the economic downturn might have exacerbated wage differentials. This situation is not uncommon in most other EU countries, where the average gross yearly wages of women were 16.3% lower than those of men. Nevertheless, this rate varies widely across countries.

Finally, in this thesis, occupational mismatches, understood as mismatches in education and skills, are analysed from a comparative perspective. These mismatches play an important role in determining the expected wage return of workers. According to European Commission data, almost 30% of EU workers report being overeducated, 17% being overskilled, and 15% having an excess of education and skills simultaneously. Females are more likely than males to be severely mismatched (Flisi et al., 2014) [60] and, in that sense, occupational

mismatches might be helpful to understand unexplained differences in the gender wage gaps that are not justified by the productive characteristics of workers. As explained above, this analysis will be performed both for Spain and for a set of other countries.

1.6 Hypotheses

Having outlined the research objectives, the state of the art and the current situation of the Spanish case, the following hypothesis can be formulated:

a Research objective 1

Low academic performance is a strong predictor of school failure and early school dropout. Therefore, the factors that have an incidence on the probability of a student obtaining a score below level-2 on the PISA test should be similar to those that predict early school dropout.

Hypothesis 1: Parents' socio-professional status and household possessions are the main determinants of academic performance.

Hypothesis 2: Immigrant students are at greater risk of school failure.

b Research objective 2

The educational decisions of individuals are also influenced by the macroeconomic environment and by the dynamics of the labour market, especially in times of economic unrest. At least two possible effects might exert opposing influences on the decision to remain or abandon the school system when unemployment is persistent: the “family" effect, which generates a negative response in the decision to remain at school when the students' parents are unemployed and the household finances are affected, motivating the student to drop out and look for a job; and, the “local labour market" effect, which prompts students to remain at school due to the perceived lack of labour opportunities in their communities.

Hypothesis 1: The situation of the labour market affects academic performance.

Hypothesis 2: There is a negative relationship between the employment rate and academic performance.

c Research objective 3

Individuals who are occupationally mismatched, particularly those who are overeducated and overskilled, may suffer a penalization in their expected return compared to those individuals who are correctly matched to their job. This penalization might vary depending on the gender of the worker. Therefore, taking the occupational mismatch of workers into consideration might prove useful for understanding a part of the unexplained wage differentials of women compared to men. If common patterns can be identified across countries, it may be possible to shed some light on the reasons behind the persistent gender wage gap.

Hypothesis 1: Overeducation and overskilling mismatches have a negative effect on wages.

Hypothesis 2: The effect of overeducation and overskilling on wages is stronger in the case of women.

Hypothesis 3: Part of the wage gap attributable to the different effect of overeducation and overskilling by gender is related to discrimination.

1.7 Methodology

The methodological strategies adopted in addressing the above research objectives are presented below.

In order to identify the effect that personal, household and school characteristic have on the risk of school failure among students in Spain (research objective 1), the basic framework employed was that of the education productional function, and information from the OECD Programme for International Student Assessment (PISA), the periodic international evaluation of the general competence of 15-year-old students. The sample covers the four PISA evaluations completed during the first decade of the new millennium: 2000, 2003, 2006 and 2009, providing information about the students, the schools they attend, and their parents. A multiple imputation strategy (Rubin, 1987) [127] was adopted for dealing with missing values. The hierarchical structure of the data was taken into account using a two-level model specification (Raudenbush and Bryk, 2002) [121], where the first level represents the students' personal and household characteristics, and the second level represents the school attributes. The dependent variable is a dichotomous variable which takes a value of one if the student is at risk of school failure and zero otherwise. The hierarchical structure of

the data was handled through a multilevel logistic regression that allows the researcher to consider the nested structure of data within larger units of aggrupation, and the influence of variables within the same level as well as across levels.

To explore the effects of labour market dynamics on the schooling decisions of young adults in Spain (research objective 2), we once again exploited the PISA databases (2006, 2009 and 2012) so as to incorporate information from before, during and after the economic crisis. The “family" and “local labour market" effects of unemployment were modelled in a hierarchical linear regression. Missing values were handled using a multiple imputation strategy. The dependent continuous variable is the student's outcome on the test of mathematical competence.

Independent regressions for each of the PISA samples were performed in order to identify how the evolution of unemployment affects schooling decisions before, during and after the economic downturn. Thereafter the trends in the performance were analysed using the methodology proposed by the OECD (2014) [111] for the periods 2006 to 2009, 2006 to 2012 and 2009 to 2012. Finally, a new regression for a pooled sample comprising all PISA waves was estimated, the aim being to enrich the analysis by incorporating the heterogeneity of the time-component variation and controlling for fixed effects.

Finally, the differences in the returns to occupational mismatches between genders (research objective 3) were studied using a number of measures of educational and skills mismatches. The input required to build these measures of mismatch was data from the Survey of Adult Skills, which forms part of the 2013 Programme for the International Assessment of Adult Competences (PIAAC). This survey provides information on the key abilities and workplace skills needed for individuals to be able to participate adequately in society. The sample corresponds to 20 OECD countries: Belgium, Chile, Czech Republic, Denmark, France, Greece, Indonesia, Israel, Italy, Japan, Lithuania, Netherlands, New Zealand, Norway, Poland, Republic of Korea, Slovakia, Slovenia, Spain and the United Kingdom. The analysis focuses on both literacy and mathematical competencies.

The effects of occupational mismatches on the gender wage gap were explored using a Mincerian-type regression, incorporating the approach reported by Verdugo and Verdugo (1989) [140] that facilitates the determination of the return to overeducation and overskilling for the whole sample and for each gender separately. Sample selection bias was controlled following a Heckman (1979) [70] approach. Subsequently, a Oaxaca-Blinder [100] decomposition (1973) was performed to reveal which part of the wage return differences between mismatched males and females can be explained in terms of productive characteristics and which part in terms of discrimination. An unconditional quantile

regression was also implemented to analyse the gender wage gap across the wage distribution. The Machado and Mata (2005) [85] technique was selected. Finally, we explored possible common patterns across countries.

1.8 Structure of the thesis

This thesis comprises four chapters, in addition to this introduction, representative of three separate studies, and the conclusions. The first study (Chapter 2) is entitled “The evolution of school failure risk during the 2000 decade in Spain". It seeks to provide an overview of the evolution of the factors that have had a significant influence on the risk of school failure by focusing on personal, household and school characteristics. A multilevel logistic regression analysis was applied to the 2000, 2003, 2006 and 2009 PISA databases. The findings show that, when considering students' reading skills, males are twice as likely as females to drop out of school, and that immigrant students are at a higher risk of dropping out than native students are. Household characteristics have a strong positive influence and so reduce the risk of school failure, while school-related variables have only a negligible effect on school failure. This study has been published in the journal“Estudios sobre Educación" of the University of Navarra in 2014.

The second study (Chapter 3), “Labour markets, academic performance and school dropout risk: Evidence for Spain" provides a better understanding of the links between labour market conditions and academic performance by disentangling the effects of unemployment. The contribution of this study is, therefore, threefold: first, it provides new evidence on the link between labour market conditions and educational decisions; second, it quantifies separately the two effects of unemployment on academic performance at age 15; third, it analyses the heterogeneous impacts of the “family" and “local labour market" effects, proxied through the unemployment rate of the school community. The analysis of the impact of unemployment on academic performance is performed through hierarchical linear regressions (HLR). Results show that academic performance at age 15 is affected by labour market conditions, and, at the same time, previous performance determines future educational decisions. Thus, these results highlight the sensitivity of students' educational decisions and academic performance to shifts in the labour market. This study has been accepted for publication in the journal “International Journal of Manpower" in 2017.

The third study (Chapter 4), “The effect of overeducation and overskill on the gender wage gap: A cross-country analysis" performs a comparative analysis of the gender wage differentials for a group of OECD countries considering the incidence of overeducation and

overqualification, and determines the extent to which differences in the gender wage gap may be attributed to educational and skills mismatches. The contribution of this study is thus twofold: first, it provides new evidence on the link between occupational mismatch and the gender wage gap; second, it explores the factors behind cross-country differences in the gender wage gap. We use Mincer wage equations, controlling for sample selection bias, in order to estimate differences by gender in the effects. We then apply Oaxaca-Blinder decompositions to determine the effects of overeducation and overskill on the gender wage gap. Finally, a Machado-Mata (2005) [85] quantile regression approach is performed at the country level. Results show that: i) overeducation has a negative effect on wages and that this effect is larger for women; ii) overskill also has a negative effect on wages, but differences by gender are less evident; and, iii) most of the gender wage gap is explained by the different returns by gender of the productive characteristics of workers, occupational mismatch being one of the factors that increases this gap. This study is in the process of being submitted as an original research article to Economics of Education Review or the Journal of Labor Economics.

Finally, chapter 5 presents the main conclusions of this thesis. It seeks to highlight the unifying thread that provides coherence for the three studies that make up this thesis and it also reports our policy recommendations and identifies future research lines.

Chapter 2

The evolution of school failure risk during the 2000 decade in Spain

2.1 Introduction

Overcoming high rates of school failure is one of the most complex challenges faced by Spanish society. National and international studies, such as Fernández Enguita et al.(2010) [58], and OECD reports (2010 [107], 2011 [108]), address the problem of the low performance of Spanish students compared to their European peers. Results are clear: Spain faces higher levels of school failure and early school dropout.

The broad definition of school failure includes all forms of not achieving the educational objectives determined by society as the minimum necessary to be integrated into the labour market and to become a productive member of the community. Accordingly, the definition of school failure chosen in this paper includes all the individuals who are not able to complete compulsory secondary education (ESO) at the age of 16.

School failure in Spain appears to have structural characteristics, as it has been present in the educational system for more than 30 years, with figures fluctuating around 30% during the last two decades. Failure to accomplish the objective proposed by the Lisbon Strategy in 2010 and difficulties in achieving the Europe 2020 strategy of reducing the early school leaver rate in EU countries to less than 10% are indicators of issues in addressing the causes of the problem.

The aim of this chapter is to analyze the evolution of the factors that determined school failure risk during the 2000 decade in Spain. In this paper, following the work of authors such as Schleicher (2007) [133], school failure risk is defined as the probability of obtaining a score below level-2 in reading competency in the Programme for International Student

Table 2.1 Risk of school failure by competencies in the Spain PISA tests

PISA waves	2000	2003	2006	2009
School failure	26.6	28.7	30.8	25.9
Reading competence	16.7	18.5	19.1	20.0
Mathematics competence	25.0	20.4	17.2	21.8
Scientific knowledge competence	20.6	20.3	14.0	17.7

Source: OECD-PISA 2000, 2003, 2006 and 2009 databases and Ministerio de Educación, Cultura y Deporte (2011).

Assessment (PISA). The selection of reading competency as the main area in this analysis is due to the emphasis that this competency has in two of the four PISA tests (2000 and 2009).

The analysis is performed using 2000, 2003, 2006 and 2009 PISA micro-data for Spain. This should allow the observation of variations in the determinants over time, and their importance as predictors of school failure risk, broadening the scope of previous works such as Calero et al. (2010) [27] or Choi and Calero (2011) [41] and permitting the introduction of methodological improvements.

Table 2.1 compares real school failure rates in Spain and the risk of school failure in all the PISA test competencies. The measure of school failure risk in PISA tends to underestimate the real volume of students who fail.

This chapter is structured as follows: Section 2.2 explains and justifies the individual, household and school-level variables considered in the analysis. Section 2.3 presents the data and the methodology, while Section 2.4 discusses the results. Finally the main conclusions are presented in the last section.

2.2 Determinants of academic performance and school failure

The present section has been divided into three sub-sections according to the three blocks of explicative variables considered in this chapter: personal, family and school characteristics.

2.2.1 Personal characteristics

The differences in the academic performance of students depend on a number of characteristics that are distinctive at an individual level, and have a direct influence on the probability of school success. Gender appears to be an important personal determinant that

affects the academic performance of students. For example, there is a significant difference in the ESO graduation rates between males and females. Using year 2006-2007 data, Fernández Enguita (2010) [58] found a variation of almost 14 percentage points for females over males in graduation rates. A capacity for organization, discipline and attention appears to be a more common characteristic of female students, who also seem to have an advantage in the learning processes. It is therefore to be expected that a positive relation will be found between obtaining a high score in the PISA reading test and being female. It is also anticipated, that a lower proportion of female students will be at risk of school failure compared with their male peers.

It is important to mention that this result does not hold for the PISA math test (Calero and Choi, 2010) [27]. Carrington et al. (2008) [34] attempted to analyze whether the gender gap in performance could be explained by teacher gender, in a study of 11-year-old British students. However they found no evidence of differences in the attitudes or approaches of teachers related to the gender of their students.

Another characteristic linked to the degree of maturity of the pupils is the month of birth of the student. Crawford et al. (2011) [47] found evidence that students born between January and March seem to obtain consistently better grades than students who were born from October to December. Similarly, García Montalvo (2011) [63], using the TIMSS and PIRLS databases for Spain, found evidence of a positive relationship between being born in the first and second quarters of the year and educational achievement measured by the scores of the students in the PISA international standardized tests. Consequently, a variable that accounts for the month of birth has been introduced to test for differences in the possible outcomes of students in the PISA tests.

Finally, students who are significantly older than their own school cohort seem to have some disadvantages in motivation, engagement and performance compared to their peers in the appropriate grade for their age (Martin, 2009) [88]. There are significant examples in the literature that provide evidence of a negative relation between grade retention and educational outcomes (Holmes, 1989 [71]; Jimmerson, 2001 [74]). However this variable has been intentionally excluded from our estimations due to endogeneity issues between academic achievement and grade repetition[1]. Table 2.2 shows the proportion of students who obtain a low result in the PISA reading competency in the four PISA evaluations.

[1]In preliminary estimations of the results presented in Section 2.4 the variable “grade retention" was included, the results suggesting a strong link between grade retention and academic performance. However, we chose not to incorporate the variable in the final estimations in order to avoid bias generated by the introduction of an endogenous variable.

Table 2.2 Probabilities of obtaining a grade under level-2 in the PISA reading test according to personal characteristics

Personal variables	2000	2003	2006	2009
Grade				
2° ESO	74.19%	68.06%	69.35%	66.96%
3° ESO	42.05%	39.32%	38.37%	35.78%
4° ESO	5.46%	9.68%	7.90%	7.38%
Gender				
Male	20.81%	24.90%	25.14%	24.19%
Female	10.94%	11.83%	13.00%	13.89%
Birth month				
January to march	14.68%	17.59%	16.47%	16.87%
October to December	19.13%	21.17%	22.01%	20.43%
Country of birth				
Spain	15.34%	17.39%	17.81%	16.98%
Others	29.95%	34.34%	36.82%	35.65%
Origin of the student				
National students	16.31%	17.85%	17.81%	16.98%
First generation immigrants	33.33%	40.27%	40.86%	38.23%
Second generation immigrants		29.54%	32.99%	35.45%
Aggregate mean	17.37%	19.95%	20.44%	19.98%

Source: Own compilation with OECD-PISA 2000, 2003, 2006 and 2009 databases

Not all the students in the sample were born in Spain. An increasing number of students among the four PISA waves used in this chapter have diverse nationalities and origins. Being an immigrant is an important characteristic that seems to increase the risk of school failure (Table 2.2) and is related with adaptation issues, such as language and cultural differences (McCarthy, 1998) [90]. A dummy variable that distinguishes those students born in Spain from those born outside the country is introduced in the 2000 and 2003 waves. In the 2006 and 2009 waves, two dummy variables are introduced to measure the effect of being either a first or second-generation immigrant on the risk of school failure, compared to native students. The language spoken at home appears to be an important determinant of the process of adaptation of immigrants to their new country (Entorf and Minoiu, 2005) [53].

2.2.2 Household characteristics

Household attributes and material resources are two important aspects to be considered in the analysis of academic performance. The Coleman Report (1966) [44] provided evidence that family background is the main factor in student academic performance.

Hanushek (1997) [67] showed that differences in household environments, such as students living in single-parent families or coming from lower socio-economical backgrounds, are relevant for individual academic achievement. Haveman and Wolfe (1995) [69] state that it is a household's background characteristics that have the most powerful effect on the academic achievement of students.

There is an important difference in the academic performance of students whose parents belong to a managerial or professional category compared to those from families where the parents are manual workers (Cohen, 1987) [43]. The former are a small segment of the population and have significant advantages in school achievement, grades and completion rates compared to the latter. Consequently the model presented in Section 2.3 includes a variable that describes the household member in the highest socio-economic category.

There is ample literature that shows the relationship between the educational level of the parents and the performance of students. Ferguson et al. (1996) [57] posited that parental education accounted for about 24 percent of the variance in student's test scores; Reynolds and Temple (1998) [123] affirm that the level of education of the parents is positively associated with test scores and negatively with grade retention. Consequently it is to be expected that an inverse relation will be observed between the level of education attained by parents and the probability of students being at risk of school failure.

Table 2.3 illustrates the percentage of students at risk of school failure according to the evaluation results in the PISA reading competency and considering family attributes.

Table 2.3 Probability of obtaining a grade under level-2 in the PISA reading test considering household characteristics

Household variables	2000	2003	2006	2009
Occupation of the parents				
Working	15.59%	16.77%	15.86%	15.32%
Not working	19.12%	22.03%	24.15%	22.37%
Socio-economical category				
Skilled white-collar worker	2.91%	6.54%	8.36%	6.37%
Unskilled white-collar worker	8.03%	11.15%	11.87%	9.56%
Blue-collar worker	19.11%	20.78%	22.31%	22.05%
Parental education level				
Tertiary education	8.59%	10.51%	12.43%	10.84%
Compulsory secondary education (ESO)	11.17%	16.87%	16.74%	18.43%
Primary education	24.91%	25.42%	34.89%	30.47%
Did not finish primary education (ESO)	45.26%	40.95%	46.33%	55.52%
Home educational resources				
Computer, calculator, books and dictionary	12.23%	14.80%	14.06%	13.56%
Does not have these resources	22.53%	27.01%	23.63%	22.04%
Household cultural possessions				
Literature, poetry and works of art	10.71%	12.37%	13.88%	13.37%
Does not have these possessions	27.37%	26.50%	29.05%	29.55%
Aggregate mean	17.37%	19.95%	20.44%	19.98%

Source: Own compilation with OECD-PISA 2000, 2003, 2006 and 2009 databases

The possession of cultural resources is expected to have a negative relation to the risk of school failure (Berger et al., 2005) [18]. Similarly the possession of a large number of books is correlated with early reading competence in individuals (Aikens and Barbarin, 2008) [2]. Two variables that account for the household's cultural environment and the specific material possessions within the household are incorporated in the model, and it is predicted that they will reduce the probability of obtaining a score below level-2 in the PISA reading test.

2.2.3 School characteristics

School-level determinants refer to the characteristics of the schools, the type of students who enroll in them and their material resources and their allocation. The most relevant factors

affecting the risk of school failure seem to be to a significant extent already set before the students enter school, but it is important to determine if the school magnifies or reduces the differences between students with diverse characteristics and risk factors.

One significant determinant associated with the characteristics of educational institutions is school ownership. State schools contain a larger proportion of immigrant students and they have students with a wider range of characteristics and family backgrounds. The graduation rate from ESO in private schools is almost 20 percent above the rate in public institutions (Ministerio de Educación, 2009) [95].

Interaction between students sharing certain characteristics contributes to the enhancement or the reduction of the academic performance of peers (Coleman et al., 1966) [44]. Accordingly the following determinants are introduced: the proportion of females in the school population, the socio-economic characteristics of the students and also the educational level attained by parents.

Another important characteristic related to the interaction between students is their origin. We have therefore introduced a variable that measures the proportion of immigrants in the schools. Two different thresholds are used: 20% for the years 2000 and 2003, and 30% for the years 2006 and 2009. The use of two different thresholds is justified by the progressive increase in the percentage of immigrant students who entered the Spanish educational system during the 2000 decade.

Calero and Escardíbul (2007) [28], using PISA-2003, found that there is no significant difference in the performance of the students at different types of schools in Spain, and Cordero et al. (2011) [46] reach the same conclusion, demonstrating that this variable was not relevant in the academic achievement of the students tested in PISA-2009. Considering this empirical evidence, it is to be expected that differences in the ownership of the school are not relevant determinants of school failure risk if socio-economical characteristics are controlled for.

The controversial debate over the importance of a reduced teacher-student ratio in the classroom is also considered in the analysis. While authors such as Krueger (2002) [77] argued for the need to reduce class size in order to improve the quality of education, Chingos (2010) [40] and Hanushek (2003) [68] consider student-teacher ratio to be a factor that does not enhance the quality of education or academic results. To test this, we include a variable that represents student-teacher ratio.

Even though the real effect of school material resources on the academic performance of students is a matter of debate due to the results of international (Hanushek, 2003) [68] and national studies (Calero, Choi and Waisgrais, 2010) [27], variables such as school size and the

students-per-computer ratio have been included with the purpose of observing their evolution throughout the 2000 decade. However, as Lavy (2012) [80] warns, resource analysis may not be adequate if endogeneity is not addressed, that is if the fact that schools with certain profiles have higher student-teacher ratios than others is not taken into consideration. This is clearly a research area to be explored in Spain in the future.

Finally, another group of variables included in the analysis covers the participation of schools in school budget allocation and course content. According to the results observed by Calero and Waisgrais (2009) [29] , the effect of these variables is not particularly significant. However, using data from different years, it is still possible to observe interesting results regarding the effect of the autonomy of schools on school failure risk.

Table 2.4 shows the rate of students who could not achieve results equal to or above level-2 in the PISA reading competency, according to school attributes.

2.3 Data and methodology

This section is divided into two parts. The first part describes the data provided by the PISA evaluations, the method employed to gather the data and the strategy necessary to handle this particular kind of database. The second part of the section outlines the econometric technique and model best suited to fit the PISA databases.

2.3.1 Data

The database for this chapter corresponds to the OECD Programme for International Student Assessment (PISA) implemented in the late nineties as a strategy for the periodic international evaluation of the general competence of 15 year-old students. The sample in the present work covers the four PISA evaluations completed during the 2000 decade.

Table 2.5 describes the size of the sample of students and schools considered for each one of the four PISA evaluations in the present analysis.

In each PISA evaluation, emphasis is placed on one specific competency. These competencies were reading in the 2000 test, mathematics in 2003, scientific knowledge in 2006 and again reading in 2009.

The PISA test consists of three survey forms, one for the students, one for the parents and one for the head of the school. In Spain the parent questionnaire was omitted and, consequently, the main sources of information were the students and the schools. The student questionnaire is designed to gather information about specific competencies and also

Table 2.4 Probability of obtaining a grade under level-2 in the PISA reading test considering school characteristics

School variables	2000	2003	2006	2009
Size of the community where schools are located				
Community +100.000 inhabitants	12.63%	17.18%	15.29%	13.55%
Community + 1.000.000 inhabitants	10.64%	18.02%	17.45%	18.14%
Type of school				
Private government independent	4.68%	11.13%	10.14%	8.13%
Private government dependent	10.32%	13.02%	13.69%	11.69%
Public	21.06%	22.93%	23.74%	23.97%
School size (number of students)				
Over the average	13.93%	12.44%	14.29%	17.00%
Under the average	19.63%	20.29%	22.22%	20.66%
Percentage of girls				
Over the average	17.23%	17.27%	18.39%	18.61%
Under the average	14.39%	17.57%	19.86%	19.78%
Ratio students-computers				
Over the average	17.88%	22.34%	22.60%	21.28%
Under the average	14.76%	14.68%	16.91%	17.33%
Ratio students-teacher				
Over the average	8.91%	10.65%	13.81%	12.71%
Under the average	20.54%	22.93%	23.25%	23.00%
Ratio immigrants-national				
Immigrant students over 20%	33.33%	38.14%	30.89%	26.34%
Immigrant students over 30%	28.81%	53.42%	40.00%	30.05%
Educational school environment				
Tertiary education	7.19%	8.93%	10.07%	10.65%
Compulsory secondary education (ESO)	14.12%	19.23%	19.84%	22.59%
Primary education	26.16%	32.07%	30.64%	58.59%
Socio-professional school environment				
Skilled white-collar parents	0.00%	0.00%	3.50%	1.96%
Unskilled white-collar parents	4.51%	6.50%	8.86%	7.13%
Blue-collar parents	18.32%	20.69%	20.79%	20.45%
Budget management autonomy				
School responsibility	16.22%	11.33%	15.18%	9.76%
Not a school responsibility parents	25.00%	21.04%	22.63%	19.96%
Course content autonomy				
School responsibility	15.89%	17.83%	19.40%	17.57%
Not a school responsibility parents	17.72%	18.11%	18.60%	19.35%
Aggregate mean	17.37%	19.95%	20.44%	19.98%

Source: Own compilation with OECD-PISA 2000, 2003, 2006 and 2009 databases

Table 2.5 Sample size and target population of the Spain PISA tests from 2000 to 2009

	2000	2003	2006	2009
Total population of 15 years old	451,685	454,064	439,415	433,224
Students sample	6,214	10,791	19,604	25,887
Weighted students participants	399,055	344,372	381,686	387,054
Schools sample	185	383	686	889

Source: OECD-PISA 2000, 2003, 2006 and 2009 databases.

background information regarding the personal and household characteristics of the students. The school questionnaire collects information from the schools and makes it possible to match information between students and schools.

The PISA sampling is carried out in two stages. In the first stage, a sample of schools is randomly selected from a list of eligible schools. In the second stage a subsample of 35 students aged 15 years is randomly selected within each school. Student and school level sampling weights are provided to correct marginal deviation from the random probability process of selection. Consequently, sampling weights have been used throughout our analysis to avoid bias in population parameter estimates[2].

The educational outputs of the PISA evaluation are the scores in the standardized tests that give different values to the abilities of the students. Non-observable random factors that can affect the test result are controlled for through a set of plausible values. This characteristic of the PISA tests implies the need to incorporate the plausible values for each competency in the analysis. Performance in each PISA competency is measured through a set of 5 plausible values. These values cannot be individually interpreted as scores, but as a set they are accurate in describing the performance of the population, as they contain a random error variance component. Estimations must therefore be performed five times per competency and then averaged. Standard errors are calculated following the same method, and the total variance is determined adding the measurement error and the sampling variances.

Missing values, present in all the PISA evaluations, require a particular approach that allows the true nature of the data to be left unaffected. The methodology that seems most appropriate is the Multiple Imputation strategy (Rubin, 1987) [127], a procedure by which missing data are imputed several times to produce different complete data estimates of the parameters. The estimated parameters are combined to produce an overall estimate of the complete data parameters with minimal effects on the standard error. Multiple imputation

[2]OECD (2009a) [105] provides a thorough description of the use of sampling weights with PISA.

by chained equations was performed using the Stata package. This software offers a more flexible method of dealing with missing values, compared to fully-parametric methods like maximum likelihood and Bayesian analysis.

2.3.2 Methodology

The analysis of PISA data requires multilevel modeling in order to account for the hierarchical structure of the data and a logit-type specification for the binary response dependent variable. We use a two-level formulation proposed by Raudenbush and Bryk (2002) [121], the first level corresponds to data from the students clustered within schools, and the second level captures the influence of school factors.

Traditional techniques are not suited to accounting for the hierarchical and clustered structure of the data. Multilevel regression takes into account the nested distribution of the data within larger units of concentration, calculating a different equation for each level of aggregation. These models not only identify the relations of different variables within the same level but also the influence of variables from one level to another.

The first level of the multilevel regression corresponds to *i* students selected in the second stage of the PISA survey and, the second level, to *j* schools sampled in the previous stage of the survey. The logistic random intercept for a dichotomous y_{ij} dependent variable is modeled according to Raudenbush and Bryk (2002) [121].

Two level random-intercept fixed-slope logistic regression model

Level 1 model

1. $\eta_{ij} = \beta_{0j} + \sum_{k=1}^{n} \beta_{1j} X_{kij} + \xi_{ij}$

2. $\rho_{ij} = \dfrac{exp(\eta_{ij})}{1 + exp(\eta_{ij})}$

$$y_{ij} = 1 \quad \text{with probability} \quad \rho_{ij}$$
$$y_{ij} = 0 \quad \text{with probability} \quad 1 - \rho_{ij}$$

3. $\log\left(\dfrac{\rho_{ij}}{1-\rho_{ij}}\right) = \beta_{0j} + \sum_{k=1}^{n} \beta_{1j} X_{kij} + \xi_{ij}$

Level 2 model

4. $\beta_{0j} = \gamma_{00} + \sum_{l=1}^{m} \gamma_{01} Z_{lj} + u_{0j}$

5. $\beta_{0j} = \gamma_{10}$

Full model

6. $\log\left(\frac{\rho_{ij}}{1-\rho_{ij}}\right) = \gamma_{00} + \gamma_{10} X_{kij} + \gamma_{01} Z_{lj} + u_{0j} X_{kij} + u_{0j} + \xi_{ij}$

In the equations X_{kij} represents the student level covariates and Z_{lj} school level covariates. $\beta_{0j}..\beta_{1j}$ represent regression coefficients. $u_{0j} \sim N(0, \tau_{00})$ are school specific random intercepts, uncorrelated across schools and uncorrelated with covariates. $\xi_{ij} \sim$ logistic are student-specific residuals, uncorrelated across students and schools, uncorrelated with u_{0j} and with covariates. All multilevel estimations have been performed using HLM 6.25, which follows, for two-level models, the methodology suggested by Pfefferman et al (1998) [116].

2.4 Results

The regression results are shown in Table 2.6. Annex A describes the variance reduction analysis. The interpretation of the odd-ratios depends on the specification of the variables and the sign of the coefficients. When the variable has a positive coefficient, every 0.1 over 1.0 represents a 10% increase in the probability that the student scores below level-2 in reading competency. On the contrary, if the coefficient is negative, every 0.1 under 1 represents a 10% decrease in the probability of obtaining a grade under level-2.

In two particular cases, variables were replaced due to the differences in the sample size and the information available in the database for the four PISA evaluations.

In the personal characteristics, the variable in the 2000 and 2003 regressions (COB) that distinguished students born in Spain from those born outside the country, was replaced by two variables, FGIM (first generation immigrant) and SGIM (second generation immigrant), for 2006 and 2009, with the purpose of illustrating the evolution of immigrant students in the school system in Spain in greater detail.

In school characteristics, as has been explained, the chosen threshold for the ratio of immigrant students/total students in the school was increased from 20% (IRATIO20) for 2000 and 2003 to 30% (IRATIO30) for 2006 and 2009 in order to account for the progressive arrival of immigrant students in Spain and to capture significant effects.

Two mechanisms were used complementarily to check the correlations between independent variables with a correlation matrix and also to test the variance inflation factors (VIFs)[3]. Level 1 and level 2 variables were inspected separately.

Table 2.6 shows the coefficients of the two-level logistic regressions, the signs of which reflect the relation between the explanatory variables (personal, household and school characteristics) and the dependent variable (probability of obtaining a score below level-2 in reading competency), and the odd-ratios or likelihood ratios and the robust standard errors.

2.4.1 Personal variables

The first variable with considerable statistical significance in the regressions for the four years is the gender of the student. Being male increases one's chance of obtaining a result under level-2 in reading competency by over 100%. This result is consistent in all the regressions: 100% in 2000, 130% in 2003, 100% in 2006 and 110% in 2009.

This result implies that for every female student who obtains an unsatisfactory result in reading competency, there are at least two males in the same situation. The uniformity of the results over time seems to support the conclusions of Bertrand and Pan (2011) [19] that showed that females have higher rates of success in the school system.

As suggested by Crawford et al. (2011) [47], the month of birth seems to have a close relation to academic achievement and the cognitive skills of the students. This is not a surprising result as Sprietsma (2010) [136], using PISA 2003 data, found a positive long term effect for relative age when entering primary school in 10 out of 16 countries analyzed.

The negative sign of the coefficient for the variable for students who were born from January to March, suggests that students born in the first quarter of the year are less likely to obtain a deficient result in the reading competency of the PISA evaluation. The same pattern is observed for the variable that describes the students born in the last quarter of the year, but with the opposite sign, indicating that students born from October to December are more prone to fail to obtain a result over level-2 in reading competency. These results also seem to

[3] We consider that if any of the VIF values is larger than 4 there are multicollinearity problems associated with the variable.

Table 2.6 Multilevel logistic regressions fixed effects. Estimation for the probability of obtaining a score below level-2 in the reading competency PISA evaluations

	2000 HLM Coeff.	2000 HLM Odd Ratio	2003 HLM Coeff.	2003 HLM Odd Ratio	2006 HLM Coeff.	2006 HLM Odd Ratio	2009 HLM Coeff.	2009 HLM Odd Ratio
Constant	-0.981^{***}	0.4	-0.973^{**}	0.4	-0.611^{***}	0.5	-1.375^{***}	0.3
	(0.360)		(0.343)		(0.162)		(0.150)	
Gender	0.688^{***}	2.0	0.844^{**}	2.3	0.710^{***}	2.0	0.727^{***}	2.1
	(0.111)		(0.083)		(0.062)		(0.059)	
Born in the 1st quarter	-0.072	0.9	-0.045	1.0	-0.221^{***}	0.8	-0.141^{**}	0.9
	(0.140)		(0.093)		(0.081)		(0.067)	
Born in the 4th quarter	0.224^{*}	1.3	0.072	1.1	0.160^{*}	1.2	0.179^{***}	1.2
	(0.115)		(0.092)		(0.075)		(0.070)	
Skilled white-collar worker	0.287	1.3	-0.138	0.9	-0.001	1.0	-0.004	1.0
	(0.314)		(0.231)		(0.174)		(0.166)	
Non-skilled white-collar	-0.247	0.8	-0.226	0.8	-0.332^{***}	0.7	-0.378^{***}	0.7
	(0.178)		(0.129)		(0.100)		(0.100)	
Country of birth	-0.412	0.7	-0.301	0.7				
	(0.244)		(0.215)					
1st generation immigrant					0.939^{***}	2.6	0.983^{***}	2.7
					(0.148)		(0.098)	
2nd generation immigrant					-0.086	0.9	0.307	1.4
					(0.336)		(0.280)	
Occupation of the parents	0.243^{*}	1.3	-0.030	1.0	-0.395^{***}	0.7	-0.086	0.9
	(0.108)		(0.084)		(0.078)		(0.063)	
Parents highest education level	-0.054^{***}	0.9	-0.029^{*}	1.0	-0.048^{***}	1.0	-0.045^{***}	1.0
	(0.014)		(0.015)		(0.011)		(0.012)	
Home educational resources	-0.443^{***}	0.6	-0.423^{***}	0.7	-0.387^{***}	0.7	-0.399^{***}	0.7
	(0.100)		(0.084)		(0.069)		(0.072)	
Cultural possessions/family	-0.591^{***}	0.6	-0.458^{***}	0.6	-0.559^{***}	0.6	-0.601^{***}	0.5
	(0.108)		(0.085)		(0.076)		(0.064)	
Community + 100.000 inhabit.	0.023	1.0	0.155	1.2	-0.024	1.0	-0.113	0.9
	(0.189)		(0.175)		(0.133)		(0.155)	
Community + 1.000.000 inhabit.	-0.213	0.8	-0.255	0.8	-0.419^{**}	0.7	-0.513^{**}	0.6
	(0.273)		(0.288)		(0.215)		(0.235)	
Type of school: Private	-0.689	0.5	-0.114	0.9	-0.290	0.8	-0.006	1.0
	(0.436)		(0.368)		(0.273)		(0.359)	
Priv. government dependent school	-0.308	0.7	-0.148	0.9	-0.326^{*}	0.7	-0.159	0.9
	(0.328)		(0.257)		(0.176)		(0.260)	
Size of school by students	-0.090	0.9	-0.267	0.8	-0.099	0.9	0.042	1.0
	(0.204)		(0.198)		(0.125)		(0.137)	
Girls/school ratio	-0.098	0.9	-0.279^{**}	0.8	0.010	1.0	-0.028	1.0
	(0.180)		(0.138)		(0.118)		(0.110)	
Student/computer ratio	-0.225	0.8	0.277	1.3	-0.046	1.0	0.189	1.2
	(0.166)		(0.204)		(0.113)		(0.122)	
Student/teacher ratio	-0.170	0.8	0.121	0.9	-0.073	0.9	-0.212	0.8
	(0.326)		(0.305)		(0.168)		(0.227)	
Immigrant +20% population	0.836^{***}	2.3	0.917^{***}	2.5				
	(0.290)		(0.320)					
Immigrant +30% population					0.370^{*}	1.5	0.118	1.1
					(0.191)		(0.181)	
Most parents tertiary educ.	-0.449^{*}	0.6	-0.384	0.7	-0.280^{*}	0.8	-0.351^{**}	0.7
	(0.254)		(0.244)		(0.168)		(0.153)	
Most parents white-collar	-0.368	0.7	-0.572^{*}	0.6	-0.575^{***}	0.6	-0.169^{**}	0.8
	(0.321)		(0.284)		(0.246)		(0.276)	
School budget autonomy	-0.636^{*}	0.5	-0.066	0.9	-0.166	0.8	-0.497^{**}	0.6
	(0.338)		(0.304)		(0.132)		(0.234)	
School curricular content autonomy	-0.019	1.0	-0.072	0.9	0.074	1.1	-0.125	0.9
	(0.196)		(0.157)		(0.123)		(0.171)	
Observations	6,214		10,791		19,604		25,887	
Number of schools	185		383		686		889	

Note 1: *** p<0.01, ** p<0.05,* p<0.1, Standard errors in parentheses.

support the findings by Robertson (2011) [125], which show that students born in the first quarter of the year have a clear advantage in math and reading tests.

Regarding the origin of the students, a single variable that represents students who were born in Spain and whose parents are not from outside Spain was introduced for the 2000 and 2003 PISA evaluations. Not being an immigrant reduced the probability of being at risk of failure in reading competency by 30%, compared with first and second generation immigrant students.

For 2006 and 2009, two variables were incorporated. The first variable corresponds to first generation immigrants and, as was anticipated, this variable was statistically significant for both years. The odd-ratios suggest that students born outside Spain and whose parents are first generation immigrants had a 160% higher risk of obtaining poor results in the reading test in 2006 and 170% in 2009, compared to their peers born in Spain.

The second variable corresponds to second generation immigrant students, those born in Spain but whose parents are both from outside Spain. The regression indicates that the variable is statistically non-significant for 2006 and 2009. These results suggest that second generation immigrant students do not perform differently to students born in Spain whose parents are also born within the country. Second generation immigrants obtain remarkably better results in the reading tests compared to first generation immigrant students.

2.4.2 Household variables

The variables describing family characteristics are divided into two categories. The first, attempts to characterize the family within a specific socio-economic and professional group, considering the educational achievement of the parents. The second group of variables represents the underlying relationship between achievement at school and the educational material and cultural resources of the household.

Socio-economic composition and educational level of the household

The first variable in this category refers to parents with the highest socio-professional level: skilled white-collar workers. The results for this variable are statistically non-significant, so the odd-ratios results do not indicate a better performance by the students whose parents belong to this group in particular. However, students whose parents belong to the non-skilled white-collar socio-professional category seem to have a consistently lower probability of obtaining low results in the four years evaluated: 20% less during 2000 and 2003, and 30%

less during 2006 and 2009, compared to the base category of blue-collar workers. It is important to notice that these results are very significant for the 2006 and 2009 years.

The general tendency seems to be that when both parents work, there is a slight reduction in the possibility of obtaining a score under level-2 in reading competency, the 2000 regression being an exception, as its positive coefficient suggests the opposite relation. The effect seems to be consistent through time, but it is not possible to obtain a conclusive result.

Highest educational achievement by the parents has the effect of decreasing the risk of obtaining an unsatisfactory result in reading competency. Cordero, Crespo and Pedraja (2013) [45] conclude that the majority of the literature using PISA data for Spain found this recurrent relation between the educational attainment of the parents and the academic results of the students. The continuous variable indicates that the higher the level of education of the parents the lower the probability of being at risk. It is important to mention that although statistically significant, the odd-ratios do not indicate an important reduction in the probability of obtaining a result below level-2 in the test, probably showing that the impact of parental education operates indirectly through other variables

Cultural and educational resources of the household

The first variable in this category (HEDR) is an index provided by PISA that measures the possession of home educational resources. Results in Table 2.6 show that the possession of these resources is statistically very significant and decreased the probability of obtaining a low result in reading competence by 40% in 2000, and by 30% during the 2003, 2006 and 2009 regressions.

The variable that incorporates the effect of cultural possessions within the family, CULT, gathers three elements together: whether the household owns classical literature, books of poetry or works of art[4]. The results of the regressions for this variable suggest that the presence of cultural elements in the home significantly reduces the risk of obtaining a result under level-2 in reading competence by 40% for 2000, 2003 and 2006. However in 2009 cultural possessions seemingly had less of an effect on performance in comparison with that observed for the previous years.

[4]Further explanation of the construction of HEDR and CULT can be found in OECD (Technical Report PISA2000 [101]; p.225); OECD (Technical Report PISA2003 [102]; p.283); OECD (Technical Report PISA2006 [103]; p. 316) and OECD (Resultados del informe PISA2009 - Aprendiendo a Aprender V [104], p.112).

2.4.3 School variables

School variables are divided into five categories: types of schools and their location, school inputs, school composition, environment of the schools and autonomy of the school.

School characteristics

The first two variables in this category are related to the size of the town/city where the school is located. They intend to capture whether the concentration of population has an effect on the results of the students that belong to schools in a certain area.

Schools located in municipalities with populations of over a million inhabitants present a remarkable result. The signs of the coefficients are stable and negative during the four years considered, but the effect of the variable, as the evolution of the coefficients suggests, appears to increase over time (Table 2.6), indicating that students from schools located in metropolitan zones have an increasingly lower probability of being at risk.

Within this category, a second group of variables considered are those that describe the schools by the type of ownership.

According to the results obtained by Calero and Waisgrais (2009) [29], the ownership of the school appears to have a neutral effect on the probability of obtaining a deficient result in reading competency under level-2, with the sole exception of the students enrolled in private publicly-funded schools in 2006, where the risk of school failure was reduced by 30% compared to students attending state schools.

The last variable contemplated in this category describes the size of the school in terms of the number of students. As can be seen in Table 2.6, the proportion of students who perform poorly in reading competency appears to be large in schools where the number of students is under the average. However the lack of statistical significance underlines the neutral effect of the variable on the risk of school failure.

School resources

In this category two variables have been taken into consideration. The first measures the student-computer ratio. The second is the student-teacher ratio. Both variables are statistically non-significant in all the regressions. Overall, results in this category are consistent with those obtained by Calero, Choi and Waisgrais (2010) [27].

School composition

Three variables are considered in this category. The first describes the percentage of girls in the school population. The regression results indicate a positive impact on reducing the probability of school failure risk when the percentage of girls in the student population is over the average, but the estimates are only statistically significantly in 2003.

A ratio of immigrant students over 20% of the student population is statistically significant for both the 2000 and 2003 regressions (Table 2.6). The positive sign of the regression indicates that students in schools with more than 20% of immigrant students had a higher risk (130% and 150% respectively) of obtaining a result under level-2 in reading competency in 2000 and 2003.

In 2006 and 2009 the variable IRATIO20 was replaced by IRATIO30. This variable was introduced because exploratory regressions using the IRATIO20 variable showed that the results were not statistically significant for these two years. With the introduction of IRATIO30, the significance threshold of the variable was increased. This variable only considered schools in which immigrant students were over 30% of the total school population. The results show that this variable appears to be slightly significant only in 2006. These results are in line with most of the PISA-based literature for Spain, such as Cordero, Crespo and Pedraja (2013) [45], which usually finds that the proportion of immigrant students in a school has negative effects on the academic results of the native students when it exceeds the 20% threshold.

School environment

Schools with a large number of parents with ISCED 5 and 6 educational levels are compared to the rest of the schools. The results of the regressions indicate, as expected, that schools with a larger proportion of parents with tertiary education have a student population that appears to have a lower risk of obtaining a result below level-2 in reading competency compared to schools where the educational achievement of the parents is lower.

The other variable in this category indicates schools in which a majority of families belong to the white-collar socio-professional category. The regression coefficients are negative and statistically significant in 2003 and 2006. The odd-ratios corroborate the idea that schools where there are more families whose parents are white-collar workers are characterized by students that appear to have a smaller probability of obtaining a poor result in the reading test evaluation. Students in this group of schools were 30% less at risk in 2000, 40% in 2003 and 2006 and 20% in 2009. As can be observed, the tendency seems to be decreasing over time.

School autonomy

Two variables measure the degree of autonomy of the school in this category. The first variable describes the level of school budgetary autonomy. The results of the regression are statistically significant in 2000 and 2009, at a level of 50% and 40% respectively, suggesting that a high degree of independence in the allocation of resources has an important effect in decreasing the risk of students obtaining low results in reading competency compared to schools with less autonomy.

The second variable denotes schools with a significant degree of autonomy in the selection of curriculum content. The lack of statistical significance implies that the variable has a neutral effect on the probability of obtaining a result under level-2 in reading competency. The regression outcomes for both school autonomy variables do not offer a conclusive result and further research is suggested for the future.

2.4.4 Trends over the decade

To compare the performance of students in reading competence over the 2000 decade, this final section describes the trends for a set of specific variables, tracking their evolution in PISA between 2000 and 2009. The calculation of changes in a variable between the 2000 and 2009 PISA tests will only be meaningful if its definition does not change in time. This is true for reading competency, the only competency for which the theoretical framework has remained unchanged throughout all the PISA tests (OECD, 2009a) [105]. We therefore calculate trend indicators following the three-step method suggested by the OECD (2009a) [105] for two outcome variables: reading performance and the proportion of students below level-2.

The results in table 2.7 suggest that performance in reading competency by Spanish students has decreased during the 2000 decade. We have therefore split the analysis by gender, family origin and school ownership in order to provide possible explanations for this fact. Among all these subgroups, we only find a significant decrease in the performance of male students. A possible explanation for this result might be that, during the 2000 decade, non-skilled workers easily found well-paid jobs in the Spanish labor market, mainly in the real estate and service sectors. In this context, a number of young –mainly male- students might not have found incentives for continuing with their studies.

The results in table 2.8, which presents the trend followed by the main outcome measured in this study, the proportion of students performing below level-2 in PISA, are consistent with this view. According to these results, the proportion of students who obtained a score

Table 2.7 Trends in PISA reading performance variables (2000 - 2009)

	2000		2009		2009 - 2000		
	mean score 2000	s.e.	mean score 2009	s.e	difference 2009 - 2000	s.e.	standardized difference
Female	505	2.8	496	2.2	−9	6.09	−1.48
Male	481	3.4	467	2.2	−14	6.38	−2.19
Female	493	2.7	481	2.0	−12	5.97	−2.01
Public schools	478	3.5	469	2.3	−9	6.47	−1.39
Private schools	515	5.1	505	3.8	−10	8.05	−1.24
Native students	494	2.6	488	2.0	−6	5.93	−1.01
1st generation immigrants	450	15.9	428	3.9	−22	17.10	−1.29
2nd generation immigrants	460	17.8	464	8.4	4	20.29	0.20

Source: OECD-PISA 2000 and 2009 databases.

Table 2.8 Percentage of students with scores below level-2 in reading competency

	2000		2009		2009 - 2000		
	% of the population	s.e.	% of the population	s.e	difference 2009 - 2000	s.e.	standardized difference
Scores below level-2 (female)	11.5%	1.1	14.6%	0.9	3.1%	1.42	0.02
Scores below level-2 (male)	20.4%	1.4	24.4%	1.0	4.0%	1.72	0.02
Scores below level-2 (total)	16.3%	1.1	19.6%	0.9	3.3%	1.42	0.02

Source: OECD-PISA 2000 and 2009 databases.

below level-2 between 2000 and 2009 remained unchanged, suggesting a lack of significant advances in the reduction of school failure risk throughout the decade.

Additionally, there are two PISA indexes, relevant to this study, that satisfy the necessary conditions for performing adequate time-trends estimations, as their definition remained identical between 2000 and 2009: HISEI (Highest International Socio Economic Index), and ESCS (Index of Economic, Social and Cultural Status)[5].

The results in table 2.9 suggest that the HISEI index, constructed around a self-reported parental occupation indicator, has remained unchanged during the 2000 decade. The ESCS index is statistically different from zero between the two samples. The trend followed by

[5]OECD (2009b) [106] describes the items included in both indexes.

Table 2.9 Trends in PISA variables other than performance: HISEI and ESCS (2000 – 2009)

	2000		2009		2009 - 2000		
	HISEI 2000	s.e.	HISEI 2009	s.e	difference 2009 - 2000	s.e.	standardized difference
Highest International Socio Economic Index (HISEI)	44.99	0.6	44.40	0.5	−0.59	0.78	−0.76
Index of Economic, Social and Cultural Status (ESCS)	−0.56	0.05	−0.31	0.03	0.25	0.06	4.29

Source: OECD-PISA 2000 and 2009 databases.

this indicator -that measures the socio-economic background based on parent's education levels and occupational status and possessions at home- seems to imply an improvement in the socio-economic conditions of households in Spain during the decade.

In other words, the reading performance of Spanish students decreased during the 2000 decade although the socioeconomic situation of their households improved. However, combining this finding with those presented in tables 2.8 and 2.9, it might also be the case that the improvement in the economic situation (reflected through the ESCS status) had a negative impact by reducing the student's incentives to study. Future studies should focus on this hypothesis. Nevertheless, the reduction in school failure rates observed in Spain since 2009 –the beginning of the economic crisis- seems to reinforce the credibility of this hypothesis and, in a sense, the expression “school failure" could be seen as unfair, as school characteristics might not be the main factors explaining the decrease in reading performance observed between 2000 and 2009.

2.5 Conclusions

The most interesting contribution of this chapter to the previous literature on school failure risk is the possibility of observing the evolution over a whole decade of a group of factors that appear to have an important impact on the academic performance of students in Spain.

Among the personal variables the gender of the students has a strong effect on the probability of being at risk of school failure. Girls consistently perform better in reading competency than males. The ratio 2 to 1 appears unvarying throughout the decade, indicating that male students have twice the probability of school failure risk compared to female students. Although the real importance of gender may be overestimated due to the specific

selection of reading competence, school failure seems to be a mainly masculine issue in Spain. As has been explained, the relationship between school failure and the labor market deserves further research.

There is also a significant difference in the results for immigrant students compared to students born in Spain. This divergence is remarkably accentuated in the risk of school failure for first generation immigrants. The accumulation of a number of pre-conditions that seem to be characteristic of these particular students suggests that immigrant students begin from a situation of enormous disadvantage compared to national students. In this sense, policies that help to ease the process of integration of immigrants into Spanish society, and policies that increase the instruction time of immigrants in the schools, could have a positive impact on the academic performance of these students.

Household characteristics offer an important insight into the influence of the family environment on academic performance. Students with highly educated parents, belonging to the white-collar socio-professional category, along with the possession of cultural and educational resources at home, show that these things have a strong positive influence on the reduction of the risk of school failure. Therefore, further efforts should be made to reduce differences in family background and educational inputs at home. Focused grants and programs could be useful tools in achieving this objective.

Students in schools where over 20% of the total school population are immigrant students and in schools with a predominance of blue-collar families face a greater risk of school failure. In schools where parents with tertiary education are predominant, the risk to their students does not seem to decrease. Policies guaranteeing a more homogeneous distribution of immigrant students among schools, keeping ratios below 20%, could have a powerful effect on the students academic performance. The strong impact of school composition variables on academic outcome seems to support the need to reduce the segregation of students between public and private publicly-funded schools. If the educational system is meant to provide equality of opportunity to young citizens, strategies should be studied for reducing the over-concentration of low socio-economic level, immigrant students and blue-collar families in public schools.

Among the school variables the location of the school in a large city has a positive influence on the reduction of school failure risk, perhaps because of the availability of social, cultural and educational resources in more densely populated areas. After controlling for the socio-economic characteristics of the schools, the students in private and private publicly-funded educational institutions do not appear to have a lower risk of school failure than students from state schools.

The evolution of the determinants during the ten years covered by the study shows particularly stable behavior over time for personal characteristics which become, in general, more significant during the last half of the decade. Most of the household variables are also exceptionally stable during these ten years. Parental occupation deserves a special mention, appearing positively related to the risk of school failure in 2000 and then negatively related for the rest of the years. A possible hypothesis is that there is not necessarily a reduction in the risk of school failure for the economically active households, but perhaps there is an increase in the risk for those families facing problems in the labor market due to the economic crisis.

In contrast to the first-level variables, most of the school determinants are not significant during the decade. However, this fact does not imply the nonexistence of a relation between these variables and school failure risk. One reason behind the non-significance of school resources could be the specification of the model, focused on measuring the mean effects on the population sample. A possible alternative for future research is the analysis of heterogeneous effects by considering different subgroups.

It must be acknowledged that the cross-sectional analysis of the study of school failure risk is a limitation of this research. This chapter, however constrained by these restrictions, has still provided relevant information on the cumulative processes that surround school failure. Indeed it has been shown that while the proportion of students at risk of school failure has remained relatively unchanged during the 2000-2009 period, overall performance has decreased. Although unable to establish a causal relationship, our study suggests the existence of factors external to the educational system, such as the labor market situation and the sectorial structure of Spain, which may be affecting these trends.

Future studies should simultaneously include the outcomes for the three competencies and, ideally, should draw on panel data. Also, more detailed studies going more deeply into the question of immigrant students and the difficulty of measuring peer effects would be relevant in subsequent works.

Chapter 3

Labour markets, academic performance and school dropout risk: Evidence for Spain

3.1 Introduction

Labour market dynamics and the expectations of finding a job are believed to be strong determinants of individuals' educational decisions (see Belzil and Leonardi, 2007 [17]; Raaum and Røed, 2006 [119]). Indeed, labour market conditions can exercise two separate effects on schooling decisions. On the one hand, unemployment can increase the risk of school dropout and undermine academic performance via a first channel that we will refer to as the “family" effect of unemployment. Parental unemployment can lead to a deterioration in household conditions and in educational resources. In extreme cases, it may even compel young adults to complement the household income (Montmarquette et al., 2007) [98]. On the other hand, unemployment may have a positive influence on the individual demand for education via a “local labour market" effect. As labour opportunities worsen, the opportunity cost of education falls, and individuals are encouraged to stay in school (Albert, 2000) [3].

Additionally, both types of unemployment effect (“family" and “local labour market") may be heterogeneous. For example, students from lower socio-economic groups might be more sensitive to labour market conditions (Ehrenberg and Brewer, 1994) [52]. At the same time, male students may be more responsive to changes in the labour market, particularly in societies where differences in gender stereotypes are accentuated.

The aim of this chapter is to provide a better understanding of the links between labour market conditions and academic performance by disentangling the aforementioned effects of

unemployment. The contribution of this study is, therefore, threefold: first, it provides new evidence on the link between labour market conditions and educational decisions; second, it quantifies separately the two separate effects of unemployment on academic performance at age 15; third, it analyses heterogeneous effects of the “family" and “local labour market" – proxied through the unemployment rate of the school community – effects.

We address these research questions by focusing on the case of Spain, which constitutes a particularly interesting case study. Spain has recently experienced a period of intense economic growth and development (1996-2007), during which the construction and service sectors boosted labour opportunities for young adults and low-skilled workers (Arranz and García-Serrano, 2012) [10] – between 2000 and 2007, the general rate of unemployment fell from 10.6 to 8.3%. This was followed by an economic downturn, so that by 2008 the adult unemployment rate had climbed back up to 11.3%. Thereafter, the situation deteriorated rapidly and the general unemployment rate reached 24.8% in 2012.

The early school dropout rate maintained a consistent trend at around 30% from the end of the 1990s until 2008. Since 2000, this rate seems to have increased during periods of economic growth and to have fallen during periods of economic crisis (Figure 3.1). In the years between 2008 and 2012, the early school dropout rate recorded a marked shift falling from 38 to 28.8% for males, and from 25.7 to 20.8% for females (INE, 2015) [72]. Yet, in 2014, Spain still had the highest early school dropout rate in the European Union (21.9%). In this regard, Figure 3.1 seems to point to the negative relationship between unemployment and school dropout rates as suggested by Lassibille and Navarro (2008) [79].

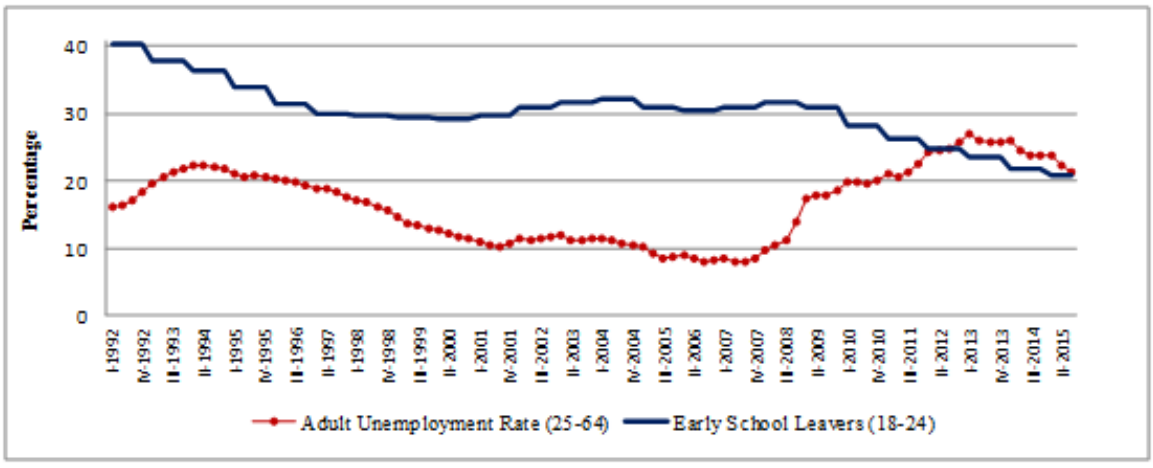

Fig. 3.1 Adult unemployment rate vs. early school leavers (18 to 24-year-olds)
Source: EUROSTAT and Spanish Labor Force Survey (EPA). Quarterly data.

We explore the “family" and “local labour market" effects of unemployment on the academic performance of Spanish students at age 15, using linear multilevel models applied to microdata from the OECD's Programme for International Student Assessment (PISA). Fifteen is a critical age in the Spanish education system, as compulsory education finishes at sixteen – also the legal working age. Given the fact that early school leaving decisions are usually the product of a cumulative process and tend to be preceded by a fall in academic performance (Barrington and Hendricks, 1989) [14], focusing on 15-year-old students seems an appropriate choice for addressing our research questions for the Spanish case.

Our results show that: i) parental unemployment has, at most, a mild effect on academic performance; ii) the “school community" unemployment effect reduces performance, although this effect decreases as general unemployment rates rise; and iii) the “school community" unemployment effect is stronger for students with a low socio-economic profile.

The remainder of the chapter is structured as follows: First, we briefly review the literature which assesses the relationship between labor market dynamics and individuals' schooling decisions. Next, we present the data and the methodology. We then turn to describe the results of the analyses. Finally, the main conclusions are presented in the last section.

3.2 Literature review

Better educated workers have higher activity rates and face a lower risk of unemployment (Gautier, 2002 [64]; Stern et al., 1989 [137]). However, during economic crises, unemployment rates grow and the monetary returns to education might fall, affecting significantly the population with the lowest levels of formal instruction (Black et al., 2005) [21]. Therefore, schooling decisions may be closely related to changes in the labour market and their impact on the family and community. Indeed, Defloor et al. (2015) [48] conclude that the quality of the first job depends considerably on individual effort, but that this effort depends largely on such circumstances as educational attainment and the situation of the labour market.

Several studies have provided evidence of the impact of changes in labour markets on academic performance and school dropout. Concerning the “family" effect of unemployment, family income may be a strong predictor of educational attainment (Carneiro and Heckman, 2002 [33]; Entwisle et al., 2005 [54]). Therefore, if the economic conditions of the household deteriorate, academic performance may be affected through a number of channels: firstly, via a reduction in the educational resources of the household (Strauss and Thomas, 1995) [138] and the socio-economic characteristics of the students (Maani and

Kalb, 2007 [84]; Rees and Mocan, 1997 [122]) and ; secondly, via a deterioration in the household environment (Rumberger, 2011) [128]; and, thirdly, via the fact that a loss of household resources may prompt young adults to leave school and seek employment. Additionally, higher unemployment rates can be perceived as an indicator of negative future returns to the investment in education. Thus, an increase in the adult unemployment rate may increase the probability of future expected unemployment, which may in turn reduce the returns to education, and therefore discourage students from enrolling in secondary education (Micklewright, Pearson and Smith, 1990) [96].

Concerning the “local labour market" effect of unemployment, Bickel (1989) [20] and McNeal (1997) [94] establish a negative relation between labor market conditions and early school dropout decisions. Rees and Mocan (1997) [122] and Rivkin (1995) [124] argue that there is a negative relationship between the general rate of unemployment and the proportion of high school students dropping out from school in any given year. Black, McKinnish and Sanders (2005) [21] show that an increase in the demand for low-skilled workers may increase the opportunity costs of education, discouraging the acquisition of further schooling. This argument is similar to those defended by Aparicio (2010) [9] and Malley and Muscatelli (1999) [86] . Finally, Duryea and Arends-Kuenning (2003) [51], drawing on data from urban Brazilian students, found that an increase in local labour market opportunities made children more likely to leave school.

Another major strand in the literature has been unable to detect significant evidence of a real impact of labour markets on individuals' schooling decisions. Manski and Wise (1983) [87], drawing on data from the National Longitudinal Study of the Class of 1972 in the United States, found only weak evidence to support an interaction between local labour market opportunities and school permanency. In addition, Card and Lemieux (2001) [32], exploring return to college trends in the United States, the United Kingdom and Canada, found that state level unemployment has no effect on educational attainment. Likewise, Warren and Lee (2003) [141], using data from the 1988 National Educational Longitudinal Study and the 1990 United States Census, conclude that labour market conditions have no significant effect on high school dropout rates. Finally, Schady (2004) [132] concludes that the macroeconomic crisis in Peru had no significant effect on students' school attendance rates, but found a significant increase in the mean educational attainment of those young adults affected by the crisis compared to those who were unaffected.

In the case of Spain, the effect of the relationship between the economic cycle and school dropout decisions has been investigated by Peraita and Pastor (2000) [113] and Petrongolo and San Segundo (2002) [115]. These studies reveal an opposite relationship between the level of income, parental education and labour market conditions on the decision to drop

out of school during primary and secondary education. Additionally, Aparicio (2010) [9] explored the recent construction boom in Spain and its impact on school dropout rates via changes in labour conditions, concluding that this temporary shock affected more significantly the employment perspectives and educational decisions of low-educated males.

Finally, the relationship between gender and unemployment seems particularly marked in the case of women living in societies in which the expectation is for them to take on more traditional roles. In Spain, for example, there continues to be a high incidence of males in the traditional role of breadwinner (Cebrián and Moreno, 2013) [35]. As such, it is likely that the schooling decisions of males will be more sensitive to changes in the labour market.

3.3 Methodological approach and data

Our study draws on information from the 2006, 2009 and 2012 PISA waves, which allows us to monitor the academic performance of 15-year-old students both before and during the economic crisis. The analysis of the impact of unemployment on academic performance is undertaken using hierarchical linear regressions (HLR).

The basic framework of the analysis is the education production function. The dependent variable is the student outcome (measured by PISA) as a function of the characteristics of the individuals, households and schools. The structure of the education production function is shown in equation (1), which considers HLR:

$$y_{ij} = \gamma_{00} + \sum_{k=1}^{\rho} \beta_{k0} X_{kij} + u_{0j} + \sum_{k=1}^{\rho} u_{kj} X_{kij} + \xi_{0j} \tag{3.1}$$

where y_{ij} are the five plausible levels of the mathematical competencies of students participating in the respective PISA wave. We focus specifically on mathematical competencies as they are an important determinant of educational choices (Aughinbaugh, 2012 [12]; Falch, Nyhus, and Strøm, 2014 [55]). Ananat et al. (2011) [6] suggest that in conditions of economic unrest and high unemployment, mathematics scores are most likely to fall. In the long run, mathematical competencies are closely linked to labor market outcomes (Checchi and Werfhorst, 2014 [38]; Joensen and Nielsen, 2009 [75]; McIntosh and Vignoles, 2001 [93]).

X_{kij} are a set of variables for individual i enrolled at school j; u_{0j} is the disturbance term at the individual level; u_{kj} is the error term for the schools, measuring the deviation of each

school from the mean slope; and finally ξ_{ij} corresponds to a series of random or stochastic effects for the HLR model.

We monitored the evolution of the unemployment-related variables during the period under analysis using two strategies. First, a regression was estimated for each of the PISA waves considered (2006, 2009 and 2012). We then tracked the evolution in performance considering three periods: 2006–2009, 2006–2012 and, finally, 2009–2012. As such, we calculated trend indicators following the three-step method suggested by the OECD (2014) [111]. More specifically, we analyzed the evolution of the mathematical competencies for four subgroups. This exercise should provide deeper insights to both labour market effects.

Second, we estimated a pooled sample across the three waves in order to exploit the heterogeneity of the time-component variation. Both the individual and the grouped regressions are controlled by fixed effects, including dummies for each of the regions, and also year dummies in the case of the pooled regression.

In both strategies (i.e. the cross-sectional and pooled samples), the personal and household explanatory (level-1) variables included are the following: gender; an index of economic, social and cultural status (ESCS), and student origin (national students and first and second generation immigrant students). The independent variables related to school characteristics (level-2) are: the size of the community in which the school is located; type of school (publicly owned, privately owned, or concertado – privately owned but government dependent); the percentage of girls by school; the student-teacher ratio; the average ESCS by school; the ratio of immigrant students by school; the degree of autonomy regarding the school's budgetary allocation and in determining the school's curriculum content[1].

We identify the “family" effect of unemployment as those households in which both parents are unemployed. Table 3.1 shows that when a student's father or mother is jobless, they are more likely to obtain a deficient result on the PISA test, and consequently are at a higher risk of abandoning the school system.

Identifying the “school community" unemployment effect with PISA is not straightforward. PISA data do not provide information on the students' place of residence[2]. Thus, we take an alternative approach. Two variables are incorporated in order to measure research question 2, namely, the “school community" unemployment effect: 1) the average

[1]Summary statistics of the explanatory variables are available upon request.

[2]PISA only provides information on the region (Comunidad Autónoma) and the size of the town/city/village where the school is located. Preliminary analyses linking regional unemployment rates from the Spanish Labour Force Survey were performed (results available upon request). However, unemployment rates may vary greatly within regions.

Table 3.1 Percentage of Spanish students under level-2 in the PISA mathematical competencies assessment, by parental labour status, 2006 to 2012

Personal and household variables	2006	2009	2012
Labour status of the father			
Unemployed	26.76%	33.32%	30.93%
Occupied	16.90%	20.35%	17.53%
Labour status of the mother			
Unemployed	21.95%	28.29%	24.28%
Occupied	15.76%	18.60%	17.65%
Household occupation status			
Both parents unemployed	30.79%	42.46%	37.22%
At least one parent working	17.08%	20.77%	18.61%
Aggregate mean	17.31%	21.84%	19.73%

Source: Own compilation with OECD-PISA databases.

labour status of the fathers by school, and 2) the average labour status of the mothers by school. These two variables measure the percentage of unemployed fathers and unemployed mothers at each school. These variables are expected to capture the reality of the labour market situation closest to the students, i.e. that related to their peers' families. This assumption can be made due to the characteristics of the Spanish system for allocating students. Students in Spain are typically allocated to public and *concertado* schools using what is known as the Boston mechanism (Calsamiglia and Güell, 2014) [30].

The Boston mechanism assigns all applicants to their first-choice school, but in cases of oversubscription for a particular school, any ties have to be resolved. In the Spanish case, the criteria used to resolve all ties are primarily distance to the school from the student's home and/or parents' workplace. Applicants not accepted by their first-choice school can then opt for any free places in other schools, but only after everyone else's first choice has been honored. As such, the mechanism drastically reduces a student's chances of being accepted in a given school once they have been rejected by their first choice.

Given the system in operation, the vast majority of students are enrolled in schools that are close to their home or to their parents' workplace. Therefore, on average, the occupational status of the student's parents at each school provides an adequate proxy of the "school community" unemployment effect. However, we acknowledge that this variable is not fully

exogenous – and, so, we cannot refer to it as a “local labour market" variable. Moreover, the system used in allocating students to schools also hampers identification of the “school community" unemployment effects: if geographical segregation reproduces socio-economic segregation, it may be harder to disentangle the effect of local labour markets from the socio-economic characteristics of people participating in those labour markets. We return to this issue in the results section.

To sum up, our models analyze two effects: the “family" and the “school community" unemployment effects on academic performance. Given that the labour market situation of a household is closely related to its socio-economic characteristics, the introduction of socio-economic characteristics is needed to disentangle the effects of labor market and socio-economic variables.

Finally, in order to address our third research question – the identification of heterogeneous effects – we incorporate an interaction between household unemployment and each of the quartiles of the ESCS index. Additionally, all the models are re-estimated for two subsamples: the top and bottom quartiles of ESCS.

3.4 Empirical analysis

Results for the HLRs of each wave and the pooled sample are presented in Table 3.2. Note that the coefficients should be interpreted as associations and not as causal effects. The model tested includes all first- and second-level explanatory variables, interactions between household unemployment and each of the quartiles of the ESCS index, and also two socio-economic and cultural index variables: household ESCS and average ESCS by school.

3.4.1 Family effect

Family effect. When both parents are unemployed the academic performance of a student does not appear to worsen. However, when the interactions were excluded, the household unemployment effect was strongly significant in the 2006 regression. However, this effect became negligible as the number of unemployed households grew in later years. According to these results, the “family" effect of unemployment, which a priori was expected to have a strong influence on a student's decision to drop out of school and look for a job, does not seem to have a marked effect on a student's academic performance.

Table 3.2 Hierarchical Linear Regression: Effect of labour market, personal and school variables on mathematical competencies, 2006-2012 – Full sample

	PISA 2006 Coeff.	PISA 2009 Coeff.	PISA 2012 Coeff.	Pooled Coeff.
Intercept	489.964	477.522	482.674	491.562
	(1.49)	(1.54)	(1.26)	(0.79)
Household unemployment	−28.550	7.443	3.712	3.058
	(25.15)	(19.76)	(16.00)	(9.30)
Unemployment rate of the fathers in the school	−72.745**	−59.157**	−37.599**	−44.824***
	(36.20)	(22.02)	(18.23)	(12.83)
Unemployment rate of the mothers in the school	−26.662**	4.619	−1.553	−6.154
	(12.35)	(16.21)	(11.45)	(8.39)
Household unemployment*4th ESCS quartile	−7.983	−10.875	−1.781	−8.213
	(28.47)	(21.11)	(16.74)	(9.98)
Household unemployment*3rd ESCS quartile	4.443	−15.089	−12.769	−12.417
	(33.09)	(20.11)	(16.25)	(10.32)
Household unemployment*2nd ESCS quartile	8.451	−22.501	−22.716	−19.608
	(26.09)	(18.31)	(18.46)	(11.61)
Gender	9.509***	19.667***	16.594***	14.606***
	(1.80)	(1.74)	(1.64)	(1.01)
ESCS Index	21.373***	18.578***	11.731***	22.218***
	(1.19)	(1.02)	(4.13)	(0.61)
First generation immigrants	−34.702***	−41.309***	−32.110***	−37.988***
	(6.56)	(3.27)	(2.95)	(2.47)
Second generation immigrants	−0.110	−1.671	−1.834	−1.814
	(5.26)	(4.06)	(3.40)	(2.43)
Average ESCS by school	18.808***	13.615***	11.731***	14.167***
	(4.44)	(4.98)	(4.13)	(2.83)
Community (<100.000 inh.)	2.213	−6.784	2.020	−0.503
	(3.99)	(4.35)	(2.89)	(2.23)
Community (>1.000.000 inh.)	0.753	8.338	12.961**	8.055*
	(7.17)	(7.42)	(6.26)	(4.16)
School private public dependent	−0.011	5.647	2.766	0.637
	(8.22)	(7.84)	(5.27)	(4.75)
School public	−0.551	0.972	−3.747	−1.635
	(9.50)	(8.16)	(5.50)	(5.06)
Percentage of females by school	3.922	0.637	1.153	3.854**
	(3.04)	(3.67)	(3.17)	(1.96)
Immigrant ratio (over 30%)	−27.556**	−2.505	−21.543***	−16.919***
	(11.96)	(7.07)	(5.76)	(5.02)
Student/teacher ratio	2.301	−4.458	6.587*	3.052
	(5.71)	(4.93)	(3.45)	(2.73)
Autonomy on budget allocation	−5.514	−7.468	20.848***	5.772
	(16.89)	(6.61)	(5.36)	(6.45)
Autonomy on course content	3.158	6.547*	3.254	4.033**
	(3.35)	(3.44)	(2.83)	(1.89)
2009 year dummy				7.574***
				(2.65)
2012 year dummy				9.792***
				(2.75)
Observations	19,604	25,887	25,313	70,804
Number of schools	686	889	902	2,477

Note 1: *** p<0.01, ** p<0.05,* p<0.1, Standard errors in parentheses.

Note: The categories of reference are: Personal and household variables: being female; living in a household where at least one of the parents is working; being born in Spain. School variables: Schools with high incidence of employed fathers; Schools with high incidence of employed mothers; community where the school is located has more than 100,000 inhabitants and less than 1,000,000; privately owned schools; percentage of girls in the school; immigrant students ratio less than or equal to 29% of the school population; student-teacher ratio; school without autonomy on budget allocation and course content. The definitions of the variables included in the model are presented in Annex B.1. Goodness of fit measures presented in Annex B.2.

3.4.2 School community unemployment effect

School community unemployment effect. Contrary to expectations, the “school community" unemployment effect did not boost the academic performance of students. Attending a school in which a high proportion of the students' fathers are unemployed has a marked effect on their performance, serving, in fact, to reduce student achievement, albeit at a decreasing rate: by 73 points in 2006, 59 points in 2009, and 37 points in 2012. In the pooled sample, the effect of the percentage of unemployed fathers by school lowered student performance by 45 points. In contrast, the concentration of unemployed mothers by school was only significant in 2006, lowering the expected outcome in the students' mathematical competencies by 27 points. The “school community" unemployment effect, especially that of the local male unemployment rate, became less intense as the crisis deepened.

The “school community" unemployment effect was expected to increase student performance, given that, as labour opportunities become scarcer, students have less motivation to drop out of school and look for work. However, the results do not reveal an increase in the expected outcomes of students as unemployment spreads, although a less negative effect on academic performance is recorded as the proportion of unemployed parents rises. Nevertheless, it should be noted that this decreasing rate might result from an inaccurate estimation, given that the number of schools with a large number of unemployed parents increases as a consequence of the generalized rise in unemployment.

The effect of the local female unemployment rate proxy is not as great as that of the male unemployment rate. This result suggests the persistence of traditional gender roles in Spanish households, where the mother's labour market decisions depend on the father's labour market situation.

3.4.3 Interactions by household unemployment and ESCS index quartiles

Interactions by household unemployment and ESCS index quartiles. We modelled interactions between the “family" unemployment effect and each of the quartiles of the socio-economic index distribution. In so doing, we expect to capture heterogeneous effects. As in the earlier analyses, we estimated the models with and without the ESCS index. We found no strong evidence in support of distinctive “family" effects on the students' academic performance by their socio-economic status. However, when testing the model interactions without the ESCS variables, we found the interactions between the lowest ESCS quartile and the “family" unemployment effect to be statistically significant, at least in the 2006 regression. But,

Table 3.3 Trends in PISA mathematics performance before and after the economic downturn

	Period	Results difference	Statistically significant	Link error	Link covariance	Standard error	Standardized difference	Hypotheses testing
Change in the performance of the students	2006 - 2009	3	No	1.333		3.39	0.89	FALSE
	2006 - 2012	4	No	2.084		3.64	1.10	FALSE
	2009 - 2012	**-2**	**No**		−0.064	**5.33**	**0.37**	**FALSE**
Change in the performance of students with both parents unemployed	2006 - 2009	−19	Yes	1.333		6.15	−3.09	TRUE(-)
	2006 - 2012	−13	Yes	2.084		6.09	−2.13	TRUE(-)
	2009 - 2012	**25**	**Yes**		−0.064	**8.72**	**2.87**	**TRUE(+)**
Change in the performance of students with only father unemployed	2006 - 2009	−11	Yes	1.333		3.91	−2.82	TRUE(-)
	2006 - 2012	−9	Yes	2.084		4.10	−2.19	TRUE(-)
	2009 - 2012	**13**	**Yes**		−0.064	**5.71**	**2.28**	**TRUE(+)**
Change in the performance of students with only mother unemployed	2006 - 2009	−13	Yes	1.333		2.05	−6.33	TRUE(-)
	2006 - 2012	−5	No	2.084		2.60	−1.92	FALSE
	2009 - 2012	**21**	**Yes**		−0.064	**3.69**	**5.69**	**TRUE(+)**

Source: Own calculations based on OECD (2014) and PISA databases.

as unemployment become a more generalized phenomenon in later years, the interactions ceased to be statistically significant.

Finally, we describe the students' performance trends on the mathematical competencies test by the following subgroups: i) the whole sample of students tested; ii) students with both parents unemployed; and, students with either the iii) father or iv) mother unemployed. Results are presented in Table 3.3.

These trends are estimated using a linear transformation that equalizes new and old data. The calculation of these trend indicators employs a three-step method: First, the difference between the mean estimates of unemployment obtained from the two pairs of tests are calculated; second, the standard error of the mean estimates is obtained; and, third, the standardized difference – i.e. the ratio of the difference between the means and the difference between the standard errors – is calculated. The methodology employed in calculating these trends is explained in detail in OECD (2014) [111].

The overall performance in mathematics remained unchanged during the whole period. However, this result differs if we analyze the evolution in the performance of the different subgroups. For instance, between 2006 and 2009, there was a downturn in the performance of those students that had both parents out of work. However, the opposite is observed between 2009 and 2012. A possible explanation for this might be that, as labor opportunities deteriorated, the “family" effect of unemployment (which decreased between 2006 and 2009, and remained relatively constant between 2009 and 2012) was offset by the “school

community" unemployment effect (which increased between 2006 and 2009 but decreased between 2009 and 2012). This result holds when we consider, first, changes in the performance of students whose fathers are unemployed and, second, changes in the performance of those whose mothers are out of work.

On the whole, these trends are consistent with our hypotheses. Between 2006 and 2009, we expected to observe a decrease in mathematics test scores due to the improvement in labour opportunities for young adults and, between 2009 and 2012, a positive change in the PISA results was anticipated as a consequence of the increase in the opportunity costs of studying.

The above analyses have identified both the “family" and the “school community" unemployment effects, with the latter being statistically significant. Sensitivity analyses excluding ESCS-related variables[3], however, highlight the close relationship between unemployment and socio-economic status. For instance, the interactions between household unemployment and the ESCS quartiles are statistically significant when the ESCS-related variables are excluded from the regressions. Thus, these results suggest that the effects of unemployment may not be homogeneous across the whole distribution of ESCS. In order to test this hypothesis, we split the sample into four sub-samples, by ESCS quartiles, and proceeded to replicate the models presented in Table 3.2, for each of these sub-samples. The results for the bottom and top ESCS quartiles are shown in Tables 3.4 and 3.5, respectively.

A common finding for both subsamples is that household unemployment only seems to have a negative effect on academic performance in 2006, with the coefficients being statistically non-significant in 2009 and 2012. However, the unemployment rate of students' fathers presents different effects depending on the SES-level of the students' households: on the one hand, the performance of students in the top quartile of the ESCS distribution seems to be independent of the unemployment level of their peers' fathers; while, on the other, the performance of students in the bottom quartile of the ESCS index is negatively affected by the level of unemployment of the school community. Interestingly, the household ESCS has a greater effect on the academic performance of low-SES students (Table 3.4) than on that of high-SES students (Table 3.5). A similar conclusion can be drawn for the schools' mean ESCS levels.

[3]These analyses are available upon request.

Table 3.4 Hierarchical Linear Regression: Effect of labour market, personal and school variables on mathematical competencies, 2006-2012 - ESCS bottom quartile

	PISA 2006 Coeff.	PISA 2009 Coeff.	PISA 2012 Coeff.	Pooled Coeff.
Intercept	464.041	457.008	457.046	459.250
	(2.20)	(1.89)	(1.67)	(1.25)
Household unemployment	−84.818***	−13.192	9.330	−10.143
	(21.43)	(12.06)	(12.29)	(9.68)
Unemployment rate of the fathers in the school	−129.602***	−90.057***	−74.766***	−74.876***
	(47.82)	(30.55)	(26.32)	(18.63)
Unemployment rate of the mothers in the school	−17.314	32.994	2.080	6.537
	(17.89)	(21.64)	(16.75)	(11.67)
Household unemployment*4th ESCS quartile	73.722**	9.372	−12.521	7.399
	(24.25)	(15.63)	(14.93)	(10.90)
Household unemployment*3rd ESCS quartile	69.554**	16.385	0.667	16.995
	(31.49)	(15.06)	(15.40)	(11.76)
Household unemployment*2nd ESCS quartile	28.185	8.298	−6.376	0.906
	(40.08)	(19.01)	(14.68)	(13.22)
Gender	10.121**	22.718***	14.379***	15.760***
	(3.88)	(3.24)	(2.98)	(2.02)
ESCS Index	45.607***	33.606***	27.371***	35.798***
	(4.93)	(4.19)	(4.23)	(2.70)
First generation immigrants	−38.655***	−45.725***	−30.378***	−39.085***
	(10.12)	(5.26)	(4.77)	(3.72)
Second generation immigrants	−3.927	6.694	−7.090	−2.229
	(9.90)	(7.86)	(7.57)	(4.88)
Average ESCS by school	13.114*	20.456***	5.105	13.405***
	(6.75)	(7.29)	(6.23)	(4.36)
Community (<100.000 inh.)	0.815	2.680	1.687	3.351
	(5.37)	(6.03)	(4.15)	(3.09)
Community (>1.000.000 inh.)	−5.357	17.067	12.730	9.237
	(16.42)	(14.08)	(12.56)	(8.26)
School private public dependent	−18.625	15.711	−3.147	−0.493
	(19.62)	(12.79)	(10.32)	(11.60)
School public	1.025	11.185	−10.238	0.263
	(21.66)	(12.63)	(10.94)	(11.85)
Percentage of females by school	−1.783	−2.347	4.578	0.757
	(3.90)	(4.82)	(4.81)	(2.83)
Immigrant ratio (over 30%)	−25.213	2.480	−24.818***	−13.981**
	(16.16)	(10.87)	(8.68)	(6.76)
Student/teacher ratio	22.473**	−2.561	3.312	6.930
	(10.66)	(5.98)	(5.49)	(4.31)
Autonomy on budget allocation	10.845	−2.586	22.526**	8.814
	(24.90)	(8.81)	(10.64)	(10.25)
Autonomy on course content	10.359**	10.239**	1.924	7.529***
	(5.00)	(5.04)	(4.31)	(2.82)
2009 year dummy				11.895***
				(2.82)
2012 year dummy				12.662**
				(4.75)
Observations	4,901	6,499	6,324	17,724
Number of schools	641	834	846	2,321

Standard error in parenthesis *** p<0.01 ** p<0.05 * p<0.1
Source: Based on OECD-PISA databases.
Note: Categories of reference defined in Table 2.2. The definitions of the variables included in the model are presented in Annex Table B.1. Goodness of fit measures presented in Annex Table B.3.

Table 3.5 Hierarchical Linear Regression: Effect of labour market, personal and school variables on mathematical competencies, 2006-2012- ESCS top quartile

	PISA 2006 Coeff.	PISA 2009 Coeff.	PISA 2012 Coeff.	Pooled Coeff.
Intercept	528.277	525.494	526.460	527.154
	(2.06)	(1.99)	(1.88)	(1.29)
Household unemployment	−61.795**	−19.942	36.125	0.016
	(20.53)	(42.89)	(27.46)	(37.73)
Unemployment rate of the fathers in the school	−106.006*	−50.490	−3.842	−35.483
	(58.17)	(42.80)	(39.05)	(23.16)
Unemployment rate of the mothers in the school	−23.168	−5.507	−24.780	−16.794
	(22.88)	(22.56)	(25.91)	(14.17)
Household unemployment*4th ESCS quartile	94.662**	14.155	−41.932	1.901
	(39.53)	(43.14)	(34.67)	(38.21)
Household unemployment*3rd ESCS quartile	−3.834	46.870	−74.636**	−13.766
	(38.50)	(40.39)	(31.13)	(36.14)
Household unemployment*2nd ESCS quartile	78.438**	34.888	−22.350	21.148
	(28.43)	(35.26)	(40.29)	(39.89)
Gender	17.926**	14.285***	18.812***	17.122***
	(4.05)	(3.08)	(3.45)	(2.03)
ESCS Index	15.437***	21.092***	11.728*	15.994***
	(5.50)	(4.33)	(5.57)	(2.89)
First generation immigrants	−12.382	−29.250**	−37.574***	−26.618***
	(14.62)	(9.86)	(9.33)	(7.00)
Second generation immigrants	6.030	−3.979	−6.532	−1.395
	(8.65)	(8.41)	(5.35)	(4.34)
Average ESCS by school	21.082***	8.351	13.784	12.990***
	(6.85)	(7.81)	(8.22)	(4.46)
Community (<100.000 inh.)	2.317	−14.050*	−2.437	−4.128
	(5.90)	(6.92)	(4.45)	(3.42)
Community (>1.000.000 inh.)	6.967	3.617	11.059	6.797
	(9.68)	(9.99)	(8.67)	(5.55)
School private public dependent	3.969	−3.853	6.401	1.240
	(8.94)	(9.13)	(7.73)	(4.96)
School public	2.171	−2.685	2.904	0.623
	(10.77)	(9.96)	(7.80)	(5.78)
Percentage of females by school	11.597**	4.233	−0.376	5.691*
	(5.13)	(5.70)	(4.89)	(3.04)
Immigrant ratio (over 30%)	−22.985	7.470	−22.610*	−11.364
	(21.47)	(10.92)	(10.80)	(8.60)
Student/teacher ratio	−4.110	−1.198	11.408**	3.749
	(8.93)	(8.00)	(5.38)	(4.25)
Autonomy on budget allocation	4.401	−8.961	24.242**	8.409
	(14.61)	(9.81)	(8.48)	(6.45)
Autonomy on course content	−5.635	12.664**	2.643	3.791
	(4.67)	(6.00)	(4.49)	(3.03)
2009 year dummy				9.089**
				(4.20)
2012 year dummy				12.410***
				(4.14)
Observations	4,898	6,471	6,316	17,685
Number of schools	640	845	844	2,329

Standard error in parenthesis *** p<0.01 ** p<0.05 * p<0.1
Source: Based on OECD-PISA databases.
Note: Categories of reference defined in Table 2.2. The definitions of the variables included in the model are presented in Annex Table B.1. Goodness of fit measures presented in Annex Table B.4.

3.5 Discussion and conclusions

From an initial inspection, our results seem to suggest that household unemployment – the “family" effect –did not have a direct influence on the mathematical performance of Spanish students. However, the analyses by ESCS quartiles show that the “family" effect was very strong and statistically significant in 2006 – when unemployment rates were low. As the crisis deepened and more families from different socio-economic backgrounds became jobless, the effect of unemployment on student performance waned. In short, the socio-economic level of the unemployed in 2012 had risen compared to that of their jobless counterparts in 2006. The analysis of the trends in this performance (Table 3.3) provides relevant supplementary evidence. Thus, there was a downturn in the performance of those students with one or both parents unemployed at the beginning of the crisis (2006 to 2009), lending weight to the idea of a strong initial “family" effect of unemployment on students' academic performance. Overall these results seem to confirm our first hypothesis.

The level of unemployment of the school community has a marked negative effect on the students' mathematical performance. This is an unexpected outcome in relation to our second research question. We report a negative effect on academic performance, but its magnitude had fallen by 2012. Indeed, the effect is only relevant in the case of the students' fathers and not in that of their mothers. We recognize that these results are not sufficiently compelling to claim that the “local labour market" effect of unemployment –proxied by the level of unemployment of the school community – generated an upturn in students' academic outcomes as the crisis become more intense. This effect is nevertheless not homogeneous across the ESCS distribution: while it is strong and statistically significant for students in the bottom ESCS quartile, it is statistically non-significant for those in the top quartile. It should be stressed that these same results were obtained even after controlling for the student- and school-levels of socioeconomic status, which points to the existence of an independent effect of the level of unemployment of the school community.

As the crisis deepened (2009 to 2012), we observe a positive – or, at least, a non-negative – shift in student performance (Table 3.3). This can be partially explained by this greater reduction in the “school community unemployment" effects – compared to those attributable to the “family" effect – during the period 2009 to 2012. The results also suggest that the “school community unemployment" effects may offset the “family" effects, indicating that as the economic downturn intensified, the opportunity cost of education was lower and, therefore, students had a larger incentive to study and to remain at school. Yet, it is not possible to completely disentangle which part of the previous evolution was driven by the change in the social composition of the unemployed and which part was driven by the net

effect of unemployment on the schooling decisions of young adults. Whether this trend continued beyond 2012 needs to be addressed in future research.

Our results show that academic performance at age 15 is affected by labour market conditions, and, at the same time, previous performance determines future educational decisions. Thus, these results highlight the sensitivity of students' educational decisions and academic performance to shifts in the labor market. This suggests that strategies aimed at reducing early school dropout rates should not be restricted solely to the education system. In other words, school failure is not only dependent on schools and, hence, on education policies. Moreover, as well as showing that the business cycle affects educational performance, this paper has also revealed the greater sensitivity of students from low socio-economic backgrounds to labor market conditions. An economic crisis, thus, represents an additional hazard to students from low socio-economic backgrounds. The fact that early school dropout rates may be stemmed during economic crises – as has been reported in Spain (see Figure 3.1), due to a fall in the opportunity cost of schooling, does not mean that those students who opt to remain in education will perform well. Our results suggest the contrary: students whose parents are unemployed and that are enrolled in schools with a high incidence of unemployment tend to present a poor academic performance and, subsequently, face problems with regards to their future continuity in the educational system. This finding calls for the need to design targeted policies for students that present this profile, policies that should be more intensive during periods of economic crisis, and which eliminate dead-ends in the educational system so as to ease educational transitions.

Certain limitations in our methodological approach could have been overcome if longitudinal data had been available. Nevertheless, this paper contributes to the existing literature by providing new evidence on the relationship between short-term labour market dynamics and academic performance. More specifically, this paper represents a significant step forward in comparison to the previous literature as it has provided responses to three key questions faced by countries with high unemployment and high early school dropout. To conclude, the deep economic crisis may have had a significant effect on the perceptions of young adults, encouraging a change in their expectations and social attitudes towards education. If this were in fact the case, then we can expect to observe a continuing decline in the early school dropout rate once the economic recovery begins. As such, this may be seen as a positive, long-run effect of the crisis. However, further research will be needed to answer this question.

Chapter 4

The effects of overeducation and overskill on the gender wage gap: A cross-country analysis

4.1 Introduction

Over the last thirty years, gender wage differentials worldwide have slowly fallen (Blau and Kahn, 2016) [22]. In Europe, as in many developed economies, the magnitude of the gender wage gap varies across both countries and productive sectors, and tends to widen at the upper and lower ends of the wage distribution (Arulampalam, Booth and Bryan, 2007) [11]. Given that these differences cannot be explained solely in terms of raw discrimination and labour market institutions, alternative explanations need to be sought.

Educational attainment and acquired competencies are key determinants of wages (Quinn and Rubb, 2006) [117]. Consequently, differences in levels of education and skills, or differences in the effects of education or skills mismatches, by gender may be important drivers of the gender wage gap. Additionally, it may be the case that the labour market rewards/penalizes mismatches in education and skills differently according to gender. In this paper, we address these issues by seeking to answer the following research questions:

1. Are there gender differences in the incidence of overeducation and overqualification?

2. Do educational and skills mismatches explain the gender wage gap?

3. Are there common patterns by country?

In addressing question 1, we construct a number of indicators of occupational mismatch, in line with Flisi et al. (2014) [60] and Perry et al. (2013) [114].

Research question 2 is addressed using a Mincer (1974) [97] wage equation within a Verdugo and Verdugo (1989) [140] framework. We also implement Oaxaca-Blinder (1973) [100] decompositions to determine how much of the gender wage gap is explained by workers' endowments and how much is explained by differences in the returns to these characteristics. Additionally, we assess the homogeneity of our findings across the wage distribution using quantile regressions (Machado-Mata, 2005) [85].

Finally, we extend the analysis to the country level in order to address question 3. We perform an international comparison to identify differences and similarities in the returns to overeducation and overqualification by gender. These analyses are applied to males and females aged between 35 and 65 from 20 OECD countries, using microdata from the first and second rounds of the OECD's Programme for the International Assessment of Adult Competencies (PIAAC).

We aim to contribute to the existing literature by presenting empirical evidence for a large sample of European countries that shows the effect of educational mismatch and skills mismatch on the gender wage gap. To the best of our knowledge, this is the first paper to explore simultaneously the effect of educational and skills mismatches on the gender wage gap from a comparative perspective. Our main results indicate that: 1) females appear to present a higher incidence of overeducation than males; 2) the wage penalty affects overeducated women more significantly; and 3) occupational mismatches increase the explained and unexplained wage differentials between genders.

The remainder of this chapter is structured as follows: first, we present a brief review of the literature that assesses occupational mismatches and gender wage differentials. Next, we present the data and the methodology. We then describe the results of the analyses and draw our main conclusions in the last section.

4.2 Literature Review

At least four theories have been developed to account for educational and skills mismatches. First, human capital theory affirms that wages are dependent on an individual's investment in education (Becker, 1964) [16]. Thus, at the aggregate level, if the supply of a certain type of worker with a specific level of education increases but demand remains constant, the wages of this worker will decrease. According to this theory, the resulting mismatch between supply

and demand will affect wages temporarily but will not necessarily translate into a waste of skills. Mincer (1974) [97] develops an empirical framework approach to this theory.

Second, the signalling/screening theory argues that investment in formal education may be efficient from an individual perspective, but have no effect on the worker's productivity. Nevertheless, schooling plays a role in allocating the most productive workers to the most productive jobs. Overeducation may occur if employers use education as a screening device in the labour markets with imperfect information (Spence, 1973) [135].

Third, the job competition theory holds that potential employers use the educational level of workers as an indicator of the cost of their investments in the training of individuals to perform a specific job with asymmetric information. Employers tend to hire individuals with higher levels of education to save potential training costs. Highly educated individuals may accept a position for which they are overeducated while they compete for a better job (Thurow, 1975) [139].

Finally, labour market allocation models seek to demonstrate that the returns associated with further education depend on the quality of the allocation of heterogeneous workers to heterogeneous occupations. Therefore, the returns to human capital investments through further schooling are constrained if the available positions do not use all the workers' education (Sattinger, 1993) [131].

Educational mismatch describes the labour market misallocation between a worker's educational attainment and the educational requirements of that worker's job, and which impacts the worker's expected returns. There is a generalized consensus in the literature of the negative effect of educational mismatch on wages. Over (under) educated workers usually receive a lower (higher) wage return than that received by workers with the exact level of education demanded by the job (Chevalier, 2003 [39]; Dolton and Vignoles, 2000 [50]; Groot, 1996 [66]).

While educational mismatch has been widely explored in the literature, the effects of skills mismatch on the labour market have not been analysed in any detail. As Desjardins and Rubenson (2011) [49] point out, educational mismatch and skills mismatch, while related, do not represent the same concept. Skills mismatch, the gap between an individual's job skills and the demands of the job market, serves as another measure of labour market effectiveness, and may arise from a particular distribution of skills and from firms' preferences for particular types of skill. In this sense, measures of overeducation may not represent just how extensively the skills of an individual are used in performing a job, or how overeducated workers may not have the necessary skills to perform the job. Most of the literature suggests that skills mismatches have a negative effect on earnings, but that their effects are not as great as

those of educational mismatch (Allen and van der Velden, 2001 [4]; Green and McIntosh, 2007 [65]; Robst, 2007 [126]).

Skill differences do not fully explain the wage gaps of overeducated workers. Yet, only a small number of papers examine both sources of mismatch simultaneously. McGuiness (2003b) [91], analysing data for Northern Ireland, emphasizes the importance of differentiating between educational and skills mismatches, as they do not measure the same phenomenon. He concludes that a significant part of the wage penalty for being overeducated does not depend on skill use. Chevalier (2003) [39], after controlling for the skill heterogeneity of a sample of UK workers, finds that the overeducation effect on wages is still negative, large and significant. McGuiness and Sloane (2011) [92], using information for the UK labour market, find that being overeducated and overskilled imply a wage penalty, but overskilling has a stronger negative effect on job satisfaction. Budria and Moro-Egido (2008) [25] report evidence of skills mismatch in the Spanish labour market but only of a wage penalty in cases of strong mismatch.

A number of studies have undertaken comparative analyses of the effects of mismatch across countries. Iriondo and Pérez-Amaral (2016) [73], drawing on a rich database of European countries, find that the return on an additional year of attained education over the required education for a job is small. Santos and Sequeira (2014) [130], in an analysis of the effects of skills mismatch in a sample of 31 European countries, conclude that the most common significant result is the positive effect of underskilling on wages, while overskilling has a negative effect in just some countries. Budria and Moro-Egido (2014) [26] seek to determine the wage effect of educational mismatch across segments of the earnings distribution for a sample of twelve European countries. They conclude that the earning gap between matched and mismatched workers is not constant across the earnings distribution.

An even scarcer body of literature has focused on the incidence of mismatch and gender wage differentials. With a specific focus on wage differentials and educational mismatch, Favaro and Magrini (2008) [56] study gender wage gaps, finding that more highly educated women experience lower wage gaps than women with lower levels of education. Addabbo and Favaro (2011) [1], by contrast, report that the most significant differences in wage returns by gender are to be found among overeducated Italian females.

Salinas-Jimenez et al. (2013) [129], in an examination of the educational mismatch of males and women in Spain, conclude that the wage gap is lower for women with less education, and that the educational mismatch increases these differences. Robst (2007) [126] analyses the differences in the reasons for educational mismatch between genders, and affirms that men are more likely to be mismatched than women. While men report career related

reasons for the mismatch, women report labour market demand related reasons as the main factor behind the mismatch.

4.3 Methodological approach and data

This section is structured as follows: First (3.1), the measures of occupational mismatch are presented; second (3.2), the methodological strategy implemented to estimate the returns to educational and skills mismatches is defined; finally (3.3), a brief description of the database is provided.

4.3.1 Measurement of educational and skills mismatches

A number of instruments have been proposed for the measurement of educational and skills mismatches (Perry et al., 2014 [114]). Here, though, we focus specifically on overeducation and overskilling, given that upward mismatches have the severest consequences (Quintini, 2011a [118]). There is, however, no consensus in the literature as to which measures should be selected. We have opted to follow Flisi et al. (2014) [60] and Flisi et al. (2013) [59], but replicate our analyses using several definitions. More specifically, we build two measures of educational mismatch and four measures of skills mismatch.

The main distinction that can be drawn between educational mismatch instruments concerns their approach. Thus, here, we construct subjective (EDUC1) and objective (EDUC2) measures of educational mismatch.

EDUC1: We compare the individual's level of education and the self-reported opinion of that worker regarding the level of education needed to perform their job. If the former is higher (lower) than the latter, the individual is overeducated (undereducated).

EDUC2: We compare the worker's educational attainment in years of schooling with the average education of all individuals in the same country and in the same type of occupation. The individual is considered overeducated (undereducated) if his educational attainment is at least one standard deviation (s.d.) over (under) the average years of education for their type of occupation and country.

The measures of skills mismatch can also be constructed using subjective and objective approaches as well as by using a mixed approach. Thus, we calculate one subjective measure (SKILL1), one objective measure (SKILL2 and SKILL3), and two mixed definitions (SKILL4/5 and SKILL6/7). The measures can be defined as follows:

SKILL1: This subjective measure of skills mismatch is constructed from two variables: the first compiles information on whether the worker finds their skill use at work challenging; and the second asks workers whether they need further training at work. If individuals report that they do not use their skills to the full in their jobs, they are considered as overskilled. In contrast, if individuals claim that they need more training, they are considered as underskilled.

SKILL2 / SKILL3: SKILL2 and SKILL3 are based on the same objective definition, but applied to different competencies (literacy and numeracy, respectively). The individual's level of competence in each area is computed from a set of ten plausible values (PV), as reported by PIAAC. The average skill level is then computed for each type of occupation at the country level. The individual is considered as overskilled (underskilled) if the skill-level for each occupation in each country is at least 1 standard deviation (s.d.) higher (lower) than the average level of competencies of the workers in that occupation and country.

SKILL4 / SKILL5: This measure adheres to the mixed definition provided by the OECD (2013a [109], p.172). First, we select individuals who are well-matched according to the subjective definition SKILL1. Then, the distributions of literacy (SKILL4) and numeracy (SKILL5) skills are computed for these "well-matched" individuals according to their occupation level and country. The individual is considered as overskilled if the average of their 10 PV on the literacy (numeracy) test is higher than that corresponding to the 95th percentile of the literacy (numeracy) skill for the individuals "well-matched" for the same level of occupation and country. Individuals with a score lower than that corresponding to the 5th percentile are considered as underskilled.

SKILL6 / SKILL7: Our second mixed definition follows Allen et al. (2013) [5]. We begin by defining the level of skills as the first plausible value of each competence score in the literacy and numeracy domains. We then define the use of the skill as the mean of the scores in all the literacy or numeracy tasks. Both measures are standardized. The standardized average of the skill use of the competence is subtracted from the standardized level of the competence skill. An individual is considered as overskilled if this value is above 1.5; underskilled if the value is below -1.5; and well-matched if the resulting value is no more than 1.5 points above or below zero.

4.3.2 Measurement of the gender gap in educational and skills mismatches

Having established definitions of educational and skills mismatches, which allow us to address our first research question, we can then study their effect on the gender wage gap (research question 2). To do so, we proceed as follows: first, we estimate returns to

overeducation and overskilling using Mincerian regressions for the total sample and by splitting the sample by gender. Second, we apply Oaxaca-Blinder decompositions. Third, unconditional quantile regressions allow us to analyse the size of the gender gap across the wage distribution. Finally, we address question 3 – the existence of common patterns across countries. We outline this strategy in further detail below.

First, a Mincer wage model (eq. 1) is constructed, within the Verdugo and Verdugo (1989) [140] framework, using the following model specification:

$$\ln Y_{it} = \alpha_0 + \alpha_1 N_{it} + \alpha_2 E_{it} + \alpha_3 OV_{it} + \alpha_4 UN_{it} + X_{it}\gamma + \varepsilon_{it} \tag{4.1}$$

Where $\ln Y_{it}$ denotes the logarithm of the real gross hourly wage by hour of individual "i" at time "t". N_{it} represents the average numeracy (literacy) skills and E_{it} describes the educational attainment in years of individual i. OV_{it} is a dummy variable that takes value 1 if the individual is overeducated or overskilled. UN_{it} is a dummy variable that takes value 1 if the individual is undereducated or underskilled (zero, if the individual is perfectly matched). X_{it} is a vector of additional covariates (age, age squared and a gender dummy).

Ceteris paribus, if productivity and wages are determined by the current level of education, we should expect coefficients α_3 and α_4 to be equal to zero. However, if wages depend on the level of education required to perform a job, the number of years of education that exceed the required amount of education or skills will be unproductive and the return to these additional years of education or skills will be null. In this sense, an overeducated or overskilled worker may obtain a lower wage than that obtained by well-matched workers with similar levels of education or skills. α_3 would then be negative.

We control by sample selection bias (Heckman, 1979) [70], considering the possibility that our sample of full-time employed workers may not be randomly selected. This is especially important for the comparative analysis, as PIAAC refers to 2011/12, a year in which the economic situation varied considerably across countries. In any case, given our methodological approach, the results of our estimations should be interpreted as associations, not as causal relationships.

We then apply a Oaxaca-Blinder (1973) [100] decomposition to assess the gap in wages between males and females. Wages are decomposed into two parts: one part is explained by productivity differences between males and females; the other part can be interpreted as unexplained differences (discrimination) by gender.

The wage equation then has to be separated for males and females by applying a number of criteria: Perfectly (educated) matched workers, overeducated workers, undereducated workers, perfectly (skilled) matched workers, overskilled workers, and underskilled workers:

$$\ln(Y_{it})_M = (X_{it})_M \gamma + \varepsilon_{it} \tag{4.2}$$

$$\ln(Y_{it})_F = (X_{it})_F \gamma + \varepsilon_{it} \tag{4.3}$$

Where $\ln Y_{it}$ denotes the logarithm of the real gross hourly wage of individual "i" at time "t", X_{it} is a vector that includes another personal explanatory variables with the corresponding coefficients vector gamma, and ε_{it} is the error term which follows a normal distribution.

The wage gap is decomposed into an explained part, the difference in the mean productivities of males and females (education, skills, experience, etc.), and an unexplained part, which accounts for differences in the returns to those productive characteristics.

The difference in mean wages between males and females can be expressed as follows:

$$\overline{\ln(Y_{it})_M} - \overline{\ln(Y_{it})_F} = [(X_{it})_M - (X_{it})_F]\hat{\gamma} + [\hat{\gamma}_M - \hat{\gamma}_F](X_{it})_F + \varepsilon_{it} \tag{4.4}$$

The first term on the right-hand side of the equation represents the component of the wage differential between the mean difference of the observed explanatory variables, while the second term on the right-hand side of the equation corresponds to the unexplained wage gap component (discrimination).

Finally, we analyse the effect of overeducation and overskilling on the gender wage gap across the wage distribution. We do so by estimating Machado and Mata (2005) [85] quantile regressions. This technique decomposes the difference between male and female hourly log wage distributions into differences in the distribution of observable characteristics of the workers, and the difference in the distribution of rewards to these characteristics between males and females.

4.3.3 Data

The database is taken from the *Survey of Adult Skills*, conducted as part of the "Programme for the International Assessment of Adult Competencies" (PIAAC), published in 2013.

Developed and implemented by the OECD, this survey measures the key skills and workplace skills needed for individuals to participate effectively in society (OECD, 2013b [110]).

PIAAC measures skills in literacy, numeracy and problem solving in technology-rich environments. These competencies are measured on a scale from 0 to 500, with 10 plausible values. In this chapter we present the results for 20 OECD countries (Belgium, Chile, Czech Republic, Denmark, France, Greece, Indonesia, Israel, Italy, Japan, Lithuania, Netherlands, New Zealand, Norway, Poland, Republic of Korea, Slovakia, Slovenia, Spain and the United Kingdom) in the literacy and numeracy domains only, given that France, Italy and Spain did not participate in the problem solving assessment. Germany and Canada were omitted due to the absence of individual income data, while Russia was omitted as it contained information solely from the Moscow metropolitan area. Descriptive statistics are presented in Appendix C.

4.4 Results

In this section we present the results of the empirical analysis. The first research question (that is, the incidence of overeducation and overskilling by gender) is addressed in subsection 4.1. The second research question is addressed in two subsections: first (4.2), we assess the effect of occupational mismatches by gender; and, second (4.3), we show how educational mismatch and skills mismatches affect gender wage differentials. Finally (4.4), we analyse patterns of occupational mismatch across countries and, so, address our third research question.

4.4.1 Educational and skill mismatches across OECD countries

The percentage shares of mismatched workers are presented in Table 4.1. The different approaches to the construction of these mismatch indicators lead to marked differences in the distribution of this mismatch in the sample. While the percentage of overeducated workers is close to 25% when applying the subjective measure (EDUC1), it is only around 14% when we apply the objective measure (EDUC2). Similarly, the percentage shares of the skills mismatch measures vary according to the approach adopted in their construction; however, the incidence of overskill in the population lies between 13 and 19% for most of the instruments. Overall, females appear to present a higher incidence of overeducation than males, but across the constructed measures of skills mismatch a higher proportion of males seem to present an excess of skills. It should be borne in mind that the results vary according to the mismatch measure constructed, since these instruments are to some extent arbitrary

and, therefore, do not strictly refer to the same sub-sample. However, the percentage shares reported by these instruments are in line with those reported by Flisi et al. (2014) [60] and Perry et al. (2014) [114].

4.4.2 Occupational mismatch effects by gender

The estimations resulting from the Mincerian regressions are presented in Table 4.2. In the case of the incidence of educational mismatch across the whole sample, the two measures of overeducation indicate that workers with an excess of education face wage penalties of around 13% (EDUC1) and 15% (EDUC2), respectively. This result is in line with the literature's findings (Dolton and Vignoles, 2000 [50]; Lamo and Messina, 2010 [78]).

The same negative effect on wages is observed for males and females, but, overall, our results indicate that the wage penalty is higher for overeducated women. While the subjective measure of educational mismatch presents a negative effect of 15% on women's wages, the objective measure shows a decrease in the earnings of women of more than 20%.

As the literature suggests, undereducated workers may obtain a wage premium compared to the wage obtained by workers that are adequately matched. The subjective measure (SKILL1) indicates a premium close to 8%, while the objective measure (SKILL2) almost doubles this. Both dimensions of the measures of educational mismatch (i.e., overeducation and undereducation) are statistically significant for both genders.

When considering the skills of workers, we expect, in line with the literature, that those with a surplus (deficit) of skills will receive wage penalties (premiums) compared to the wages received by well-matched workers with the same level of skills. However, in contrast with the educational mismatch measures that clearly indicate a higher penalty for overeducated women, the skills mismatch measures are ambiguous.

As Mavromaras et al. (2011) [89] conclude, wage penalties for overeducated workers affect females more intensively, while for workers with a surplus of skills there is no clear pattern between genders. The only measures for the full sample that offer statistically significant results are those of SKILL2 to SKILL5, and SKILL7.

These five measures indicate that overskilled workers face wage penalties ranging from 3 to 10% for the full sample, which points to a wage penalty for those workers whose abilities are in excess of those required to perform their jobs (Santos and Sequeira, 2014) [130]. However, the magnitude of this effect is lower than that of the educational mismatch.

If we consider the effect of skill surplus by gender, we do not find a statistically systematic pattern in the results of the overskill measures as we do in the full sample. In fact, while

Table 4.1 Share of mismatch by definition of occupation mismatch

Variables	Full Sample	Males	Females
OEDUC1	24.70%	24.03%	25.66%
WMEDUC1	39.57%	37.98%	41.85%
UEDUC1	30.49%	33.43%	26.26%
OEDUC2	13.99%	13.95%	14.05%
WMEDUC2	71.60%	70.91%	72.59%
UEDUC2	14.41%	15.14%	13.36%
OSKILL1	54.51%	55.23%	53.48%
WMSKILL1	10.03%	9.45%	10.87%
USKILL1	35.46%	35.32%	35.65%
OSKILL2	15.79%	17.01%	14.03%
WMSKILL2	69.18%	68.16%	70.65%
USKILL2	15.03%	14.83%	15.33%
OSKILL3	16.33%	19.38%	11.92%
WMSKILL3	69.12%	67.43%	71.59%
USKILL3	14.54%	13.20%	16.49%
OSKILL4	13.10%	13.56%	12.45%
WMSKILL4	10.03%	9.45%	10.87%
USKILL4	3.98%	4.09%	3.81%
OSKILL5	13.81%	15.36%	11.58%
WMSKILL5	10.03%	9.45%	10.87%
USKILL5	3.48%	3.24%	3.81%
OSKILL6	18.27%	18.30%	18.23%
WMSKILL6	37.54%	37.65%	37.39%
USKILL6	9.94%	10.44%	9.23%
OSKILL7	19.88%	21.00%	18.26%
WMSKILL7	38.40%	39.94%	36.18%
USKILL7	9.47%	9.47%	9.45%

Notes: Full time employees between 25-65 years.

Table 4.2 Mincer regression with occupational mismatch measures corrected for sample bias selection.

Variables	Full Sample	Males	Females
OEDUC1	-0.1351^{***}	-0.1236^{***}	-0.1500^{***}
	(0.015)	(0.023)	(0.016)
UEDUC1	0.0852^{***}	0.0890^{***}	0.0693^{***}
	(0.017)	(0.020)	(0.023)
OEDUC2	-0.1545^{***}	-0.1228^{***}	-0.2072^{***}
	(0.018)	(0.024)	(0.023)
UEDUC2	0.1558^{***}	0.1346^{***}	0.1768^{***}
	(0.018)	(0.042)	(0.043)
OSKILL1	-0.0203	-0.0304	-0.0146
	(0.017)	(0.021)	(0.027)
USKILL1	0.0604^{***}	0.0567^{***}	0.0613^{***}
	(0.015)	(0.015)	(0.017)
OSKILL2	-0.0575^{***}	-0.0721^{***}	-0.0254^{**}
	(0.016)	(0.024)	(0.013)
USKILL2	0.0928^{***}	0.0697^{**}	0.1306^{***}
	(0.027)	(0.037)	(0.036)
OSKILL3	-0.1051^{***}	-0.1349^{***}	-0.0314^{**}
	(0.021)	(0.020)	(0.016)
USKILL3	0.1788^{***}	0.2107^{***}	0.1216^{*}
	(0.037)	(0.028)	(0.067)
OSKILL4	-0.0359^{**}	-0.0462^{*}	-0.0100
	(0.018)	(0.025)	(0.033)
USKILL4	0.2169^{***}	0.1769^{***}	0.2554
	(0.039)	(0.044)	(0.070)
OSKILL5	-0.0669^{**}	-0.0549^{**}	-0.0831^{**}
	(0.020)	(0.022)	(0.040)
USKILL5	0.2111^{***}	0.1750^{***}	0.2326^{***}
	(0.040)	(0.066)	(0.043)
OSKILL6	0.0171	0.0155	0.0058
	(0.016)	(0.022)	(0.023)
USKILL6	-0.0073	-0.0480^{*}	0.0681^{**}
	(0.019)	(0.028)	(0.030)
OSKILL7	0.0369^{**}	0.0435	0.0101
	(0.018)	(0.027)	(0.025)
USKILL7	0.0195	0.0056	0.0251
	(0.024)	(0.030)	(0.024)

Note 1: Own elaboration using PIAAC database. Individual sample weights and Heckman correction considered. Explanatory variables included in these estimations are: Individual numeracy skill, average years of education, a dummy variable for the individual gender, and a quadratic polynomial of age.
Note 2: Full models are available upon demand.
Note 3: Standard error in parenthesis *** p<0.01 ** p<0.05 * p<0.1

the objective measures of overskill mismatch for the literacy (SKILL2) and mathematical (SKILL3) competencies indicate that the overskill wage penalty is higher for men (7.2 and 13.5%, respectively), the mixed method proposed by the OECD (2013a) [109] for the numeracy domain (SKILL5) points in the opposite direction: that is, overskilled women face a higher penalization (8.3%) in their incomes than men.

We are, therefore, unable to establish which gender is more severely hit by the wage penalty associated with overskilling, at least for the pooled sample of countries. In other words, the coefficients and signs for the incidence of overskill vary substantially between the various measures employed.

These differences may be attributed, at least in part, to the particular nature of each of the instruments, and the way in which each measure selects a particular subsample of the population, thus yielding different percentages for those considered as overskilled.

4.4.3 Decomposition of the gender wage gap considering the incidence of mismatch

As discussed, a Oaxaca-Blinder decomposition model is implemented to observe the differences in wages by gender considering subsamples of the population that fit the different definitions of occupational mismatch. These results are presented in Table 4.3.

Oaxaca-Blinder decompositions show that the samples defined by the measures of educational mismatch and skills mismatch are statistically significant. There is a positive difference in all cases between the mean values of the hourly log income by gender, meaning that females earn lower average incomes than males.

Results for the full sample show that if females had the same productive and observable characteristics as males, their wage would be, on average, 9.1% lower. Gender differences in the productive characteristics only account for a small part of the wage gap. Therefore, most of the difference arises from differences in the returns to these productive characteristics.

If we consider the education mismatch, the difference in the average income of overeducated female workers compared to that of males in the same situation becomes even more negative. If women had the same productive characteristics as males, and both genders were overeducated, women's wages women would, on average, be 14.9% lower than men's wages, in the case of the subjective measure, and 6.5% lower, in the case of the objective instrument.

Similar results are observed for the decomposition analysis of the subsamples that fit the different definitions of skills mismatch. Women with the same endowments as men and

Table 4.3 Oaxaca-Blinder decomposition

	Wages (in log) Raw differentials	Explained Part(%)	Unexplained Part(%)	Raw wage Ratio(%)
Pooled	0.2638***	−0.0241***	0.2879***	69.81%
Sample	(0.022)	(0.004)	(0.023)	
		-9.14%	*109.14%*	
Education mismatch (subjective)/Education level – EDUC1				
Adequate	0.2706***	−0.0323***	0.3029***	76.29%
	(0.034)	(0.006)	(0.035)	
		-11.94%	*111.94%*	
Overeducated	0.3228***	−0.0481***	0.3709***	72.41%
	(0.039)	(0.007)	(0.039)	
		-14.90%	*114.90%*	
Undereducated	0.1153**	0.0066	0.1086**	89.11%
	(0.052)	(0.009)	(0.053)	
		5.72%	*94.19%*	
Education mismatch (objective)/Years of formal education – EDUC2				
Adequate	0.2341***	−0.0279***	0.2620***	79.12%
	(0.026)	(0.005)	(0.026)	
		-11.92%	*111.92%*	
Overeducated	0.5915***	−0.0382***	0.6297***	55.35%
	(0.063)	(0.014)	(0.063)	
		-6.46%	*106.46%*	
Undereducated	0.1151*	0.0236	0.0915	89.13%
	(0.061)	(0.015)	(0.062)	
		20.50%	*79.50%*	
Skill mismatch (subjective)/PIAAC self-reported specific questions – SKILL1				
Adequate	0.1261	0.0283**	0.0978	88.15%
	(0.078)	(0.012)	(0.078)	
		22.44%	*77.56%*	
Overskilled	0.2992***	−0.0475***	0.3467***	74.14%
	(0.024)	(0.005)	(0.024)	
		-15.88%	*115.88%*	
Underskilled	0.3926***	−0.0170∗∗	0.4098***	67.53%
	(0.039)	(0.017)	(0.040)	
		-4.33%	*104.33%*	

Notes: Standard error in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Table 4.3 Oaxaca-Blinder decomposition (cont.)

	Wages (in log) Raw differentials	Explained Part(%)	Unexplained Part(%)	Raw wage Ratio(%)
Skill mismatch (Objective)/RM Reading competence – SKILL2				
Adequate	0.2657***	−0.0248**	0.2904***	76.67%
	(0.027)	(0.005)	(0.027)	
		-9.33%	*109.33%*	
Overskilled	0.3764***	−0.0463***	0.4227***	68.63%
	(0.058)	(0.010)	(0.058)	
		-12.30%	*112.30%*	
Underskilled	0.0872	0.0104	0.0768	91.65%
	(0.058)	(0.014)	(0.059)	
		11.93%	*88.07%*	
Skill mismatch (Objective)/RM Mathematic competence – SKILL3				
Adequate	0.2475***	−0.0341***	0.2816***	78.07%
	(0.027)	(0.005)	(0.027)	
		-13.78%	*113.78%*	
Overskilled	0.3055***	−0.0344***	0.3400***	73.67%
	(0.061)	(0.010)	(0.061)	
		-11.26%	*111.26%*	
Underskilled	0.1465**	0.0173	0.1292**	86.37%
	(0.058)	(0.017)	(0.060)	
		11.81%	*88.19%*	
Skill mismatch (mixed method)/OECD(2013a) Reading competence – SKILL4				
Adequate	0.1261	0.0283**	0.0978	88.15%
	(0.078)	(0.123)	(0.078)	
		22.44%	*77.56%*	
Overskilled	0.4266***	−0.0327***	0.4593***	65.27%
	(0.071)	(0.011)	(0.071)	
		-7.67%	*107.67%*	
Underskilled	−0.1085	0.0135	−0.1220**	114.46%
	(0.104)	(0.027)	(0.107)	
		-12.44%	*112.44%*	

Notes: Standard error in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Table 4.3 Oaxaca-Blinder decomposition (cont.)

	Wages (in log) Raw differentials	Explained Part(%)	Unexplained Part(%)	Raw wage Ratio(%)
Skill mismatch (mixed method)/OECD(2013a) Mathematic competence – SKILL5				
Adequate	0.1261	0.0283**	0.0978	88.15%
	(0.078)	(0.012)	(0.078)	
		22.44%	*77.56%*	
Overskilled	0.4846***	−0.0229*	0.5075***	61.59%
	(0.073)	(0.014)	(0.074)	
		-4.73%	*104.73%*	
Underskilled	−0.1131	0.0666*	−0.1797	111.98%
	(0.113)	(0.037)	(0.118)	
		-58.89%	*158.89%*	
Skill mismatch (mixed method)/Allen et al.(2013) Reading competence – SKILL6				
Adequate	0.2903***	−0.0098*	0.3002***	74.80%
	(0.038)	(0.006)	(0.038)	
		-3.38%	*103.41%*	
Overskilled	0.2406***	−0.0214**	0.2620***	78.62%
	(0.044)	(0.009)	(0.044)	
		-8.89%	*108.89%*	
Underskilled	0.1158*	0.0891***	0.0267	89.07%
	(0.060)	(0.016)	(0.061)	
		76.94%	*23.06%*	
Skill mismatch (mixed method)/Allen et al.(2013) Math. competence – SKILL7				
Adequate	0.2244***	−0.0165***	0.2409***	79.90%
	(0.038)	(0.006)	(0.038)	
		-7.35%	*107.35%*	
Overskilled	0.2177***	−0.0272***	0.2449***	80.43%
	(0.042)	(0.010)	(0.042)	
		-12.49%	*112.49%*	
Underskilled	0.1823***	0.0588***	0.1235**	83.34%
	(0.052)	(0.015)	(0.054)	
		32.25%	*67.75%*	

Notes: Standard error in parenthesis *** p<0.01 ** p<0.05 * p<0.1

considered as overskilled obtain a lower income than men. For the subjective skill mismatch measure (SKILL1), female wages are 15.9% lower than males. Considering the objective skill mismatch measures for the literacy and numeracy domains (SKILL2 and SKILL3), women, on average, receive an income that is 12.3 and 11.3% lower, respectively. The OECD (2013a) [109] mixed approach to the literacy (SKILL4) and numeracy (SKILL5) competencies yields incomes that are 7.7 and 4.7% lower, respectively, for women with the same endowments as men. The Allen et al. (2013) [5] skills mismatch measures for literacy (SKILL6) and numeracy (SKILL7) present decreases in the average income of females of around 9 and 12%, respectively.

Overall, most of the gender wage gap appears to be related to differences in the returns to productive characteristics by gender (that is, in the unexplained part), whereas only a small part of gender wage differentials seem to be explained by differences in the productive characteristics of males and females (that is, the explained part of the decomposition).

When the educational and skills mismatch measures are accounted for, the explained part remains negative, indicating that female endowments are superior to those of males and that, therefore, these differences in endowments do not explain the income differences. If women had similar characteristics to those of men, the gender wage gap would increase and, consequently, the unexplained part of the wage gap would also grow, as the main driver of the income differentials between genders is the gender-differentiated return to endowments.

4.4.4 Patterns of occupational mismatch across countries

In this section, we seek to determine whether there is a correlation between the inequalities in the incorporation of women to society and the differences in returns to occupational mismatch of men and women in the labour market. This would scaffold the gender discrimination hypothesis of the unexplained part of the Oaxaca-Blinder analyses. We select just one measure of mismatch for each of the two types of mismatch explored: first, the objective measure of overeducation (EDUC2) was chosen as opposed to the indirect self-assessment instrument because this subjective measure is more vulnerable to measurement error and, therefore, more prone to bias; and, second, the mixed method for measuring skill mismatch (SKILL5), as proposed by the OECD, was selected because it combines an objective and subjective approach specifically designed for the competencies evaluated in the PIAAC database and also because it has been shown to provide fairly stable results. Here, however, we only discuss the results obtained when using the overeducation variable, given that the results from overqualification variable were strikingly similar (nevertheless, they are available in Appendix C).

Figure 4.1 plots the estimated difference in earnings for overeducated males and females against the Gender Inequality Index (GII), as reported by the UNDP (2011). As the differences in the returns to occupational mismatch by gender widen, gender inequality (as measured by the GII) also increases. Additionally, Figure 4.2 plots educational mismatch earning differences between men and women against the Global Gender Gap (GGGI), as proposed by the World Economic Forum (2011). Here, the results are very similar, though a negative relation is observed: that is, the smaller the GGGI, the larger the gender gap.

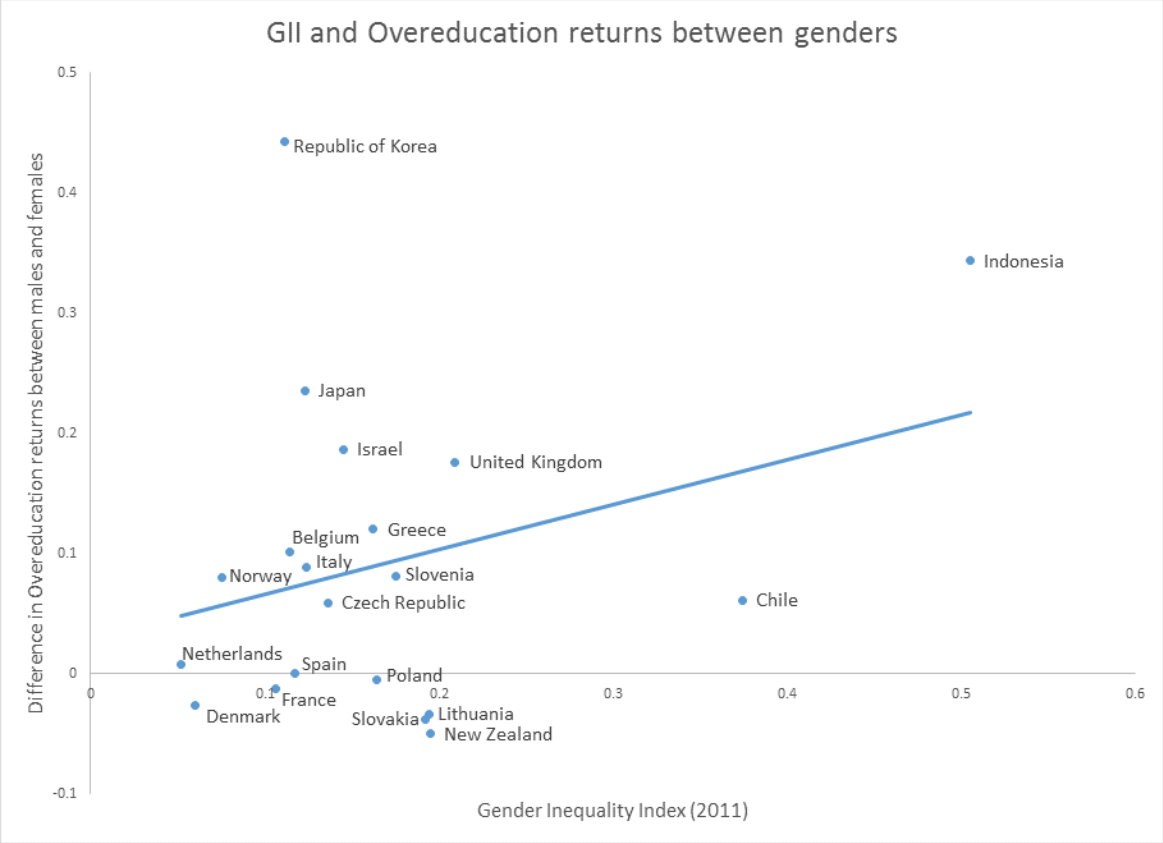

Fig. 4.1 Difference in returns to overeducation and Gender Inequality Index
Source: Own calculations and UNDP data (2011).

We established previously that occupational mismatch (understood in a broad sense to overeducation and overqualification) leads to a penalization in the returns of men and women. This penalization was higher in the case of overeducated females. A small part of this difference could be explained in terms of the workers' productive characteristics, but most of this wage gap remains unexplained, with raw discrimination apparently being the main source of these differences. Figures 4.1 and 4.2, illustrating the plots of two well-known international indexes, point in the same direction: that is, countries in which differences in

the penalization for occupational mismatch are higher are the ones in which gender inequality appears to be more prevalent.

The empowerment of women within society seems to mitigate some of the gender discrimination. As can be seen in Figure 4.3, countries with a higher percentage of females elected to parliament are negatively related and present large differences in occupational mismatch returns by gender. In countries with a high percentage of women sitting in parliament, females are more equally represented in processes of political and economic decision-making. Here, most of the countries outside Europe (i.e., Korea, Indonesia, Japan and Israel) present the strongest negative correlations between the political participation and wage returns of overeducated females.

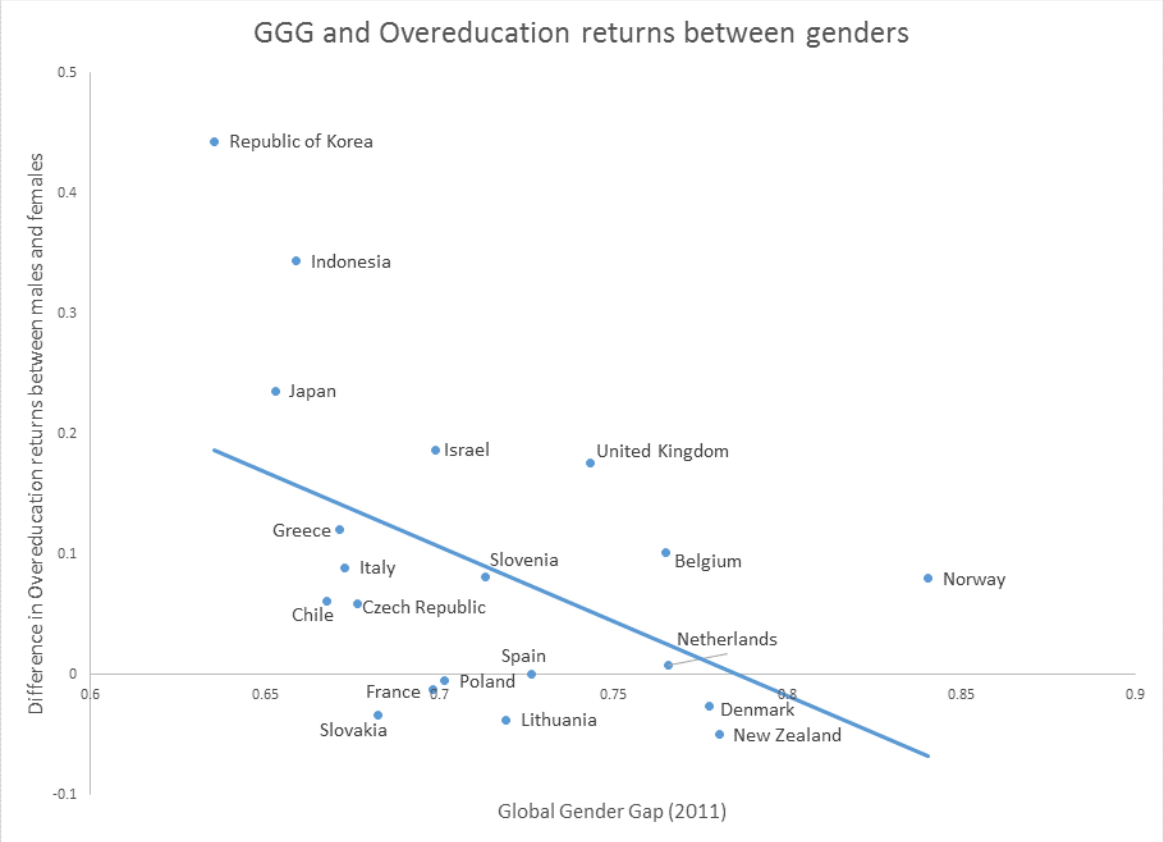

Fig. 4.2 Difference in returns to overeducation and Global Gender Gap Index
Source: Own calculations and World Economic Forum data (2011)

At the same time, female empowerment might be related to such institutions as the welfare state, and also to female participation in the labour market. More inclusive social protection schemes seem necessary to guarantee equal opportunities for both genders.

As Figure 4.4 suggests, countries with a larger share of public social expenditure present a negative relationship to marked differences in the returns to occupational mismatched workers by gender. It can be assumed that in countries in which the welfare state is more highly developed, individuals can be more independent in terms of their working conditions and personal situation. Therefore, countries in which higher numbers of unemployed workers receive periodic social security benefits are related with a lower incidence of gender inequality. Services such as early schooling and paid maternity leave are likely to be important tools for the emancipation of women in relation to the family and household. Korea, Indonesia and Israel present the clearest link between a greater (more negative) penalization of overeducated female workers and relatively low public social expenditure.

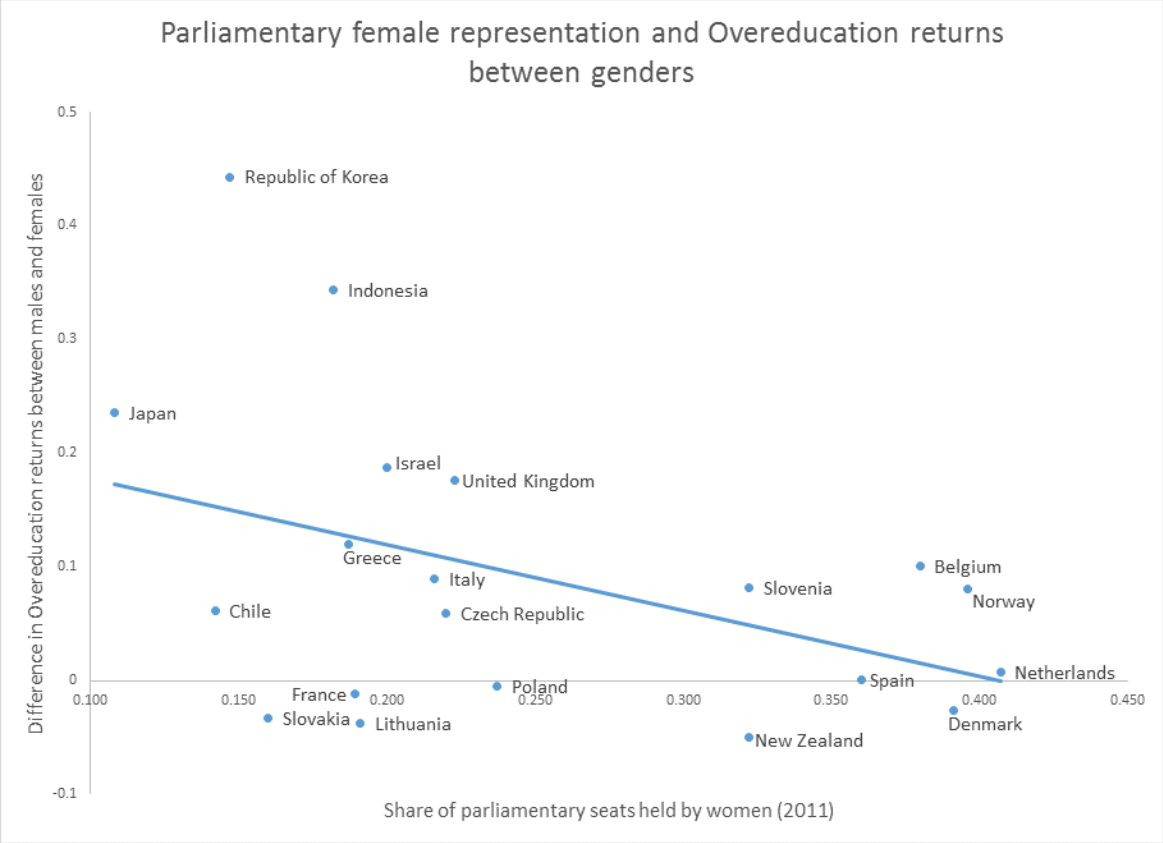

Fig. 4.3 Difference in returns to overeducation and parliamentary female representation
Source: Own calculations and World Bank data (2011)

Gender inequalities are also characterized by the lack of access of women to the workplace and to labour segregation. Figure 4.5 plots differences in the returns of occupational mismatch by gender and differences in the labour force participation of males and females. Countries with the largest gap between the labour market participation of males and females are also

those with the largest gender bias in wage returns. According to Fortin (2005) [61], anti-egalitarian views among OECD countries are strongly and negatively associated with female employment rates and the wage gap by gender. Once more Korea, Indonesia and Japan show the strongest positive correlations. In these countries, where the penalization of overeducated workers is more severe for women, males participate in the labour market more intensively than females. A similar arrangement is apparent when considering the return to overqualified workers, with Korea, Italy, Chile, Indonesia and Lithuania displaying the same positive link.

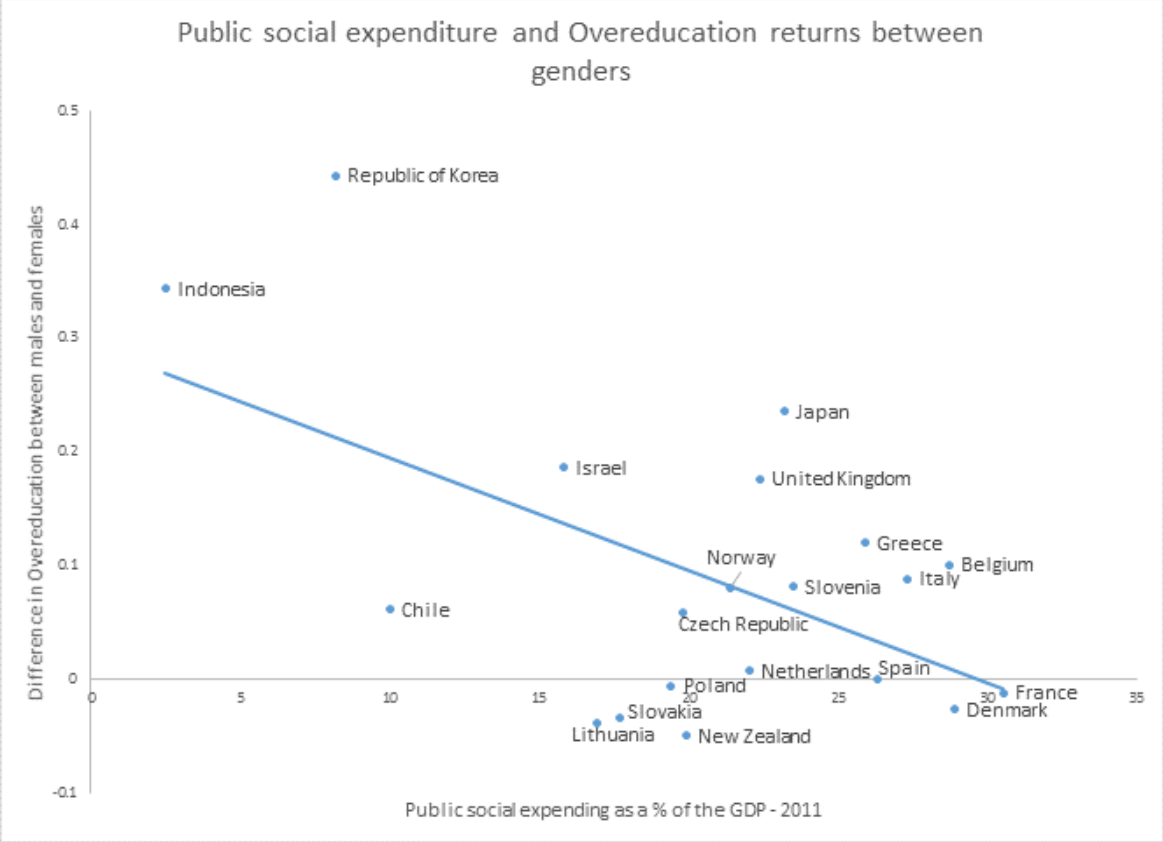

Fig. 4.4 Difference in returns to overeducation and public social expenditure
Source: Own calculations and OECD data (2011)

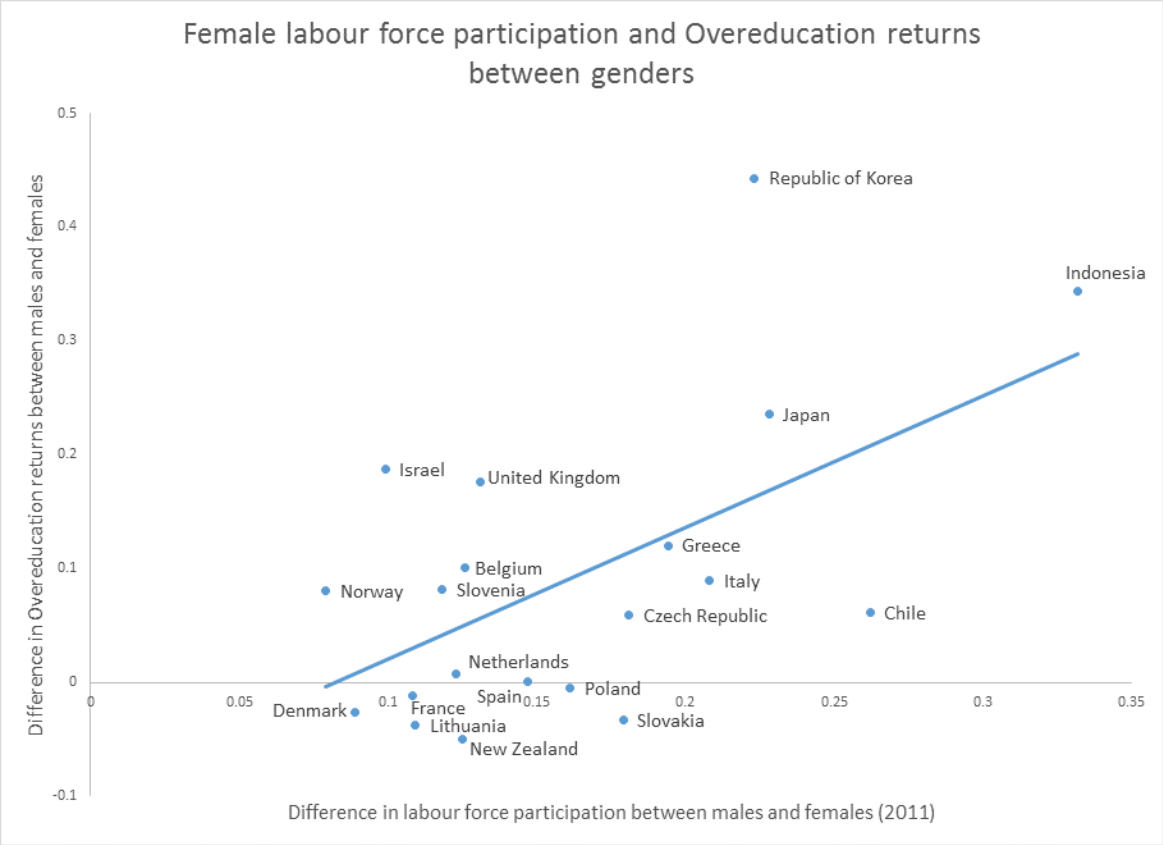

Fig. 4.5 Difference in returns to overeducation and female labour force participation
Source: Own calculations and World Bank data (2011)

4.5 Conclusions

This article set itself the aim of addressing three research questions: 1) Are there gender differences in the incidence of overeducation and overskilling? 2) Do educational and skills mismatches explain the gender wage gap? 3) Can common patterns be identified across countries? First and foremost, our results indicate that the proportion of mismatched individuals in the sample varies according to the type of mismatch measure selected and, therefore, the specific measure chosen to analyse the effect on wages is in itself a critical decision. Nevertheless, the analyses have provided consistent answers to the three questions.

First, our findings show that female earnings are lower than those of males and also that overeducation serves to penalize workers' earnings. This finding is in line with Bauer (2000) [15] and Iriondo and Pérez-Amaral (2016) [73]. In general, overeducation seems to be a condition that has a more significant effect on the monetary returns of females. Second, the wage gap seems to be greater among overeducated workers. At the same time, overskilling also has a negative effect on the wages of both genders, although the pattern is not as clear-cut as it is in the case of overeducation. These findings are coherent with those described by Allen and van der Velden (2001) [4] and by Badillo-Amador and Vila (2013) [13].

The Oaxaca-Blinder decompositions provided additional evidence of the existence of differences between the wages of males and females. Overeducation has a more severe effect on the wages of women than on those of males. However, again, when replicating this analysis using overskilling measures, our results were ambiguous.

The Oaxaca-Blinder decompositions showed that a significant part of the gender differences in the returns for occupational mismatch could not be explained by differences in the productive characteristics of individuals. Therefore, we hypothesized that these differences could be explained in terms of raw discrimination. The scatter plots provided grounds for the validity of this hypothesis. While we do not claim that occupational mismatch exacerbates gender wage inequality, we did find a close and significant association between these two phenomena. The method also allowed us to describe patterns across countries.

Asian countries, including Japan, Korea and Indonesia, presented the most extreme comparative scenarios of gender inequality and wage gap due to occupational mismatch. Other non-European countries, such as Chile and Israel, presented a similar pattern. East European countries, including Lithuania, Slovakia, Slovenia and Czech Republic (to some extent), together with Italy, presented the largest gender inequalities within Europe, while the Scandinavian countries of Denmark and Norway, along with Belgium and Netherlands, presented the most egalitarian conditions for women. This pattern seems to reinforce the

validity of the hypothesis that gender discrimination partially explains the differential effects of overeducation. However, we recognize that our analysis does not provide evidence of causal relationships. More sophisticated methods of decomposition, plus the control of unobserved heterogeneity, should help in future analyses of the dynamics of gender discrimination.

Nevertheless, this limitation and the sensitivity of our results to the different definitions of overeducation and overskilling should not be an excuse to downplay the importance of these findings. Indeed, our results not only indicate that the incidence of overeducation and overskilling are more frequent among women, but also that occupational mismatches penalize more severely the returns of the productive characteristics of overeducated women. Moreover, given the fact that part of the overeducation penalty between genders seems to be linked to raw discrimination, policies aiming at tackling overeducation and overskilling should not only focus on levelling labour demand and supply, but also on improving the institutional background so as to ensure equal opportunities in labour conditions by gender.

Chapter 5

Conclusions

The educational level and skills needed to satisfy the requirements of an increasingly demanding labour market have undergone considerable changes. Those who fail to acquire the necessary education and abilities inevitably face difficulties in finding their place in society, struggle to obtain a job, and are often underpaid. The mechanization of low-skilled jobs has worsened the work prospects of low-skilled workers. This thesis has analysed the many factors that impact the educational decisions of individuals, one of the main channels via which work skills and competencies can be acquired.

However, the educational system can at times fall short in its efforts to provide the type of skills and competencies demanded by the labour market. Often the occupational characteristics of individuals do not match the requirements of the job. The labour market tends to penalize these individuals. This situation appears to be more marked among females, who already face a persistent wage differential in comparison with their male counterparts. It might well be the case that these differences between job requirements and the education attained and the skills possessed could be useful in shedding light on the parts of the persistent gender wage gap that remains unexplained.

The results of the first paper in this thesis (Chapter 2), entitled “The evolution of school failure risk during the 2000 decade in Spain", show that the gender of the student presents a strong correlation with the risk of school failure. In fact, males underperform on reading competence tests when compared to their female counterparts. First generation immigrant students are also at a higher risk of school failure than national students, in all probability due to a number of preconditions experienced prior to their entry in the school system. A household's cultural and educational resources and the socio-professional status and educational attainment of parents are positively related to a student's academic performance and, as such, lower the risk of school failure. Students attending schools where at least 20%

of the school population is made up of immigrant students or blue-collar socio-professional households are more likely to be at risk of school failure. In these respects, it would appear necessary to overcome the segregation of schools to reverse these effects. After controlling for the schools' socio-economic characteristics, however, students at private schools and private-publicly funded schools do not appear to be at a lower risk of failure than those attending state schools.

During the period 2000-2009, the proportion of students at risk of school failure apparently remained unchanged, while the trend analysis conducted for the decade suggests that the academic performance fell slightly. As a consequence of the expansion of the service and construction sectors, low-skilled workers were able to find well-paid jobs and were, therefore, less motivated to improve their academic performance and less willing to remain in school. At the same time, the trends in the reading competence test scores and the probability of obtaining a score below level-2 on that test during that decade emphasize the lack of success of government policies and strategies aimed at reducing school failure.

The conclusions of the second research paper (Chapter 3), entitled “Labour markets, academic performance and school dropout risk: Evidence for Spain", indicate that the unemployment rate of the school community negatively affects the school performance of students as measured by their results on the mathematical competence test. This effect is only statistically significant for schools where the proportion of unemployed fathers is significantly high, becoming negligible when the unemployment rate of mothers is considered. Although we expected to find an upturn in academic performance due to the “local labour market" effect as the crisis intensified, it remained negative throughout the period analysed. Interestingly, however, its magnitude diminished, indicating a progressively positive change in academic results. The effect was not homogeneous across the ESCS distribution, affecting only those in the lower quartile. This result highlights the sensitivity of ESCS households – already at a disadvantage before the economic crisis – to changes in the economic environment. Indeed, the trend between 2009 and 2012 suggests that the “local labour market" effect might be offsetting the “family" effect. While the opportunity cost of education may increase during periods of economic unrest, and school dropout may subsequently decrease – as has occurred in Spain, our results suggest that the academic results of students whose parents are unemployed and who attend schools with a high incidence of unemployment tend to worsen and, so, their probability of remaining in the school system falls.

The results of the third study (Chapter 4), entitled “The effects of overeducation and overskill on the gender wage gap: A cross-country analysis", show that on average the earnings of females are lower than those of males, and that overeducation has a penalization effect on the expected return of workers. The gender wage gap appears to be more severe

for overeducated women. Overskill also has a negative effect on workers' expected wages, but it is not clear which gender is more severely affected by this condition. The Oaxaca-Blinder decomposition reveals that not all the differences in the wage returns of females and males can be explained in terms of the productive characteristics of individuals, and a significant part of the wage differentials is unexplained by these factors. When considering these decompositions for overeducated and overskilled workers by gender, the unexplained part in the wage gap showed a quite clear growing trend. Although we do not claim that occupational mismatch exacerbates gender inequality, we do observe a clear association between these two phenomena. This becomes more evident in the cross-country analysis. In general, overeducation and overskill are more frequent among women, and these occupational mismatches penalize more intensely the productive characteristics of females.

Having summarized the specific results of each of the three papers, it is apparent that each focuses its attention on the process of skills acquisition from different points of views and at different ages. In this regard, the three are closely related. The first and second papers provide a thorough analysis of the factors that influence the decision process of educational acquisition and an individual's permanence in the school system, first from the perspective of factors closely related to the students' environment (Chapter 2); and second, by analysing the effect of economic conditions and the labour market on the schooling decisions of individuals (Chapter 3). In the last paper (Chapter 4), we have examined how these competencies and skills are remunerated differently in the labour market, according to the quality of the match between the required occupational characteristics for the job and the occupational endowments of the individuals, and their gender. Consequently, a joint analysis of the three papers allows us to draw a number of general conclusions.

Educational outcomes, such as academic performance and school permanency, are strongly associated with factors that are determined long before a child enters the school system. Some of these are not subject to modification, such as the gender of the student, their month of birth and their country of origin, while others depend on the family background. Once in the school system, the characteristics that may adversely affect the educational process are exacerbated by factors such as the school composition and, arguably, class size. As long as the education system fails to reduce these differences, it appears education cannot function as a mechanism of social mobility and, in fact, it may even help increase or exacerbate future income and wealth inequalities. In this sense, early childhood education programs, along with public programs of health and nutrition during infancy, might be critical for reducing these differences prior to entry in the school system. However, inequity may be exacerbated if programs such as early education are not guaranteed for all students.

The socio-economic structure of education systems is also of some importance. On average, students in private schools outperform those in public schools. However, these differences disappear once the socio-economic composition of schools is taken into account. Freedom of choice in schooling can only exist if all families are guaranteed equality of access, if they are provided with information and as long as they cannot be excluded on economic grounds. This is currently not the case in Spain. A good starting point for reducing the socio-economic bias in the composition of public schools and private publicly-funded schools *(concertados)* may simply require enforcing existing legislation, which restricts the schools' capacity to select students.

The skills and competencies required by future labour markets are not readily determined. However, what is clear is that the job requirements of most economies in 15 years' time will be very different from the skills and competencies required today. Technological progress, demographic changes and industrial automation are just some of the challenges that will drive the changes in the occupational characteristics demanded by the labour force. In this future scenario, it is quite evident that those individuals who do not possess the required level of skills and competencies will have little room to participate in the workforce, and will probably end up being excluded. As manual labour becomes more obsolete and automation processes became more commonplace in working practices, individuals offering skills that are no longer required by the labour market will probably end up being left behind. Low-skilled individuals face an increasing risk of social exclusion.

In a future environment marked by profound changes in industrial structure, in which there will be a growing demand for workers with skills related to technology change and a decrease in the need for low-skilled labour, early school dropout is probably the main challenge facing the school system in Spain. The competencies and abilities required to perform the majority of jobs will probably evolve and become more complex. Consequently, individuals will not only need to update their general core skills, but also develop a set of specific abilities for their given job. In this sense, reducing early school dropout becomes a critical issue for the Spanish educational system. Some options for tackling it might well involve the removal of dead-ends in the educational system and the extension of compulsory education to the age of 18. This could comprise, for example, a combination of formal education with on-the-job training between the ages of 16 and 18. The main aim should be to help individuals develop the specific skills that are demanded by the labour market.

Given these future changes on the demand side of labour markets, it seems essential that educational systems place the emphasis on the quality, rather than on the quantity, of education. At the same time, greater coherence is needed between school systems and labour markets to ensure a mutual understanding is reached with regard to the competencies the

education system considers relevant, and the types of skills and knowledge that are necessary to perform the tasks of the workplace. In this sense, educational models aimed at reducing the degree of standardization in educational programs may be helpful to ease the transition from the school to the labour market, in order to establish coordinated mechanisms between state-controlled education and market-based demands, combined with readily available information and data to guide both education and workforce development decisions. Therefore, the way in which these competencies are exploited does not depend solely on the time individuals spend at school and on the specific abilities they acquire, but it is also heavily dependent on the instigation of appropriate coordination between economic, educational and industrial policies. Given the fact that skills are also acquired on the job or through informal channels, it becomes important to explore the development of systems that can enhance the recognition of competencies and skills for the labour market. Of course, this does not mean the formal educational system should not play a role in upgrading a worker's level of skills and competencies during his/her lifetime.

This thesis has also highlighted the existence of major inequalities in the process of the acquisition of competencies. It is not simple to address the inequalities that originate before a child enters the school system, but the first step in closing the gap between students with different backgrounds and profiles is to ensure education is a high priority on the political agenda. Targeted grant-schemes play an important role in reducing socio-economic based inequalities. At the same time, early childhood education and infant care are important for reducing inequalities before a child enters the system. However, if enrolment on early education programs and socio-economic profiles are highly correlated, then these differences are set to increase. Successful examples of the use of these policies for reducing socio-economic inequalities are provided by China, Japan and Korea (Li et al., 2017) [82], where children start early learning programs from birth until the age of six when they enter the school system. The main outcomes of these policies are the improvement recorded in individual development, the significant advance in the strengthening of individual willpower, and the positive change in emotional exchange with adults (Zhou et al., 2007) [142].

Policies that help ease the immigration process and the adaptation of foreign students should have a marked effect on the improvement of the academic outcomes of this particular group which finds itself at a high risk of school failure. Many countries run special schemes that allocate additional resources to educational institutions which cater for students who lack the minimum resources to develop their cognitive abilities or who have special needs, as is the case of many immigrant students. However, these resource allocation programs run the risk of generating more inequity when the schools that receive these extra resources end up being perceived as “marginal" schools, discouraging students, parents and teachers from

participating in them. In this scenario, an alternative would be to improve the allocation of students across schools, ensuring that a threshold of concentration of immigrant students per educational institution is not exceeded. The application of this policy will not, however, be straightforward.

The effect of labour market conditions on schooling decisions is particularly strong in the case of young adults from disadvantaged socio-economic backgrounds. Students with lower socio-economic profiles face a significantly higher risk of school failure compared to students that have enjoyed a more advantageous upbringing. Among these at-risk students, the business cycle has an important effect on academic performance and on decisions to stay on or to abandon school. It is, therefore, necessary to design policies specifically for students with this profile. One possible policy alternative might be school salary grants aimed at improving the financial situation of students from lower socio-economic backgrounds and, thus, incentivizing their permanency at school. These policies should be more intensive in periods of economic downturn in order to offer an alternative to the dead-ends that characterize the educational system and to ensure a smooth transit in the educational process during periods of economic unrest.

This thesis has also identified certain issues of gender discrimination. Gender discrimination in the labour market persists in most countries, and seems to be particularly intense in societies where the traditional role of women is associated with domestic labour and childbearing. Differences in the political representation of women, in the share of public social expenditure, and in labour force participation by gender are relevant proxies of gender inequality. Although in many countries women have caught up with men in terms of educational attainment (even exceeding them in many cases), women continue to have lower earnings than men. This gender wage gap becomes even wider for women who have an excess of education or skills for the requirements of their job.

Gender inequality in the labour market needs to be addressed by a whole series of policies instigated within an institutional framework that can help fight discrimination. While in many developed countries, women have made advances in their participation in the educational system and labour market, marked differences persist in their outcomes. Country level policies aimed at reducing occupational segregation, differentiated wages and the unequal division of paid and unpaid labour are necessary. The development of a welfare state would appear to be especially important in countries where the participation of women in the labour market is still low. Nevertheless, long-term cultural changes are needed in order to achieve gender equality in the labour market. Thus, in the meantime, measures such as mandatory participation quotas for women in all spheres of government still have a role to play.

Finally, modern societies face an increasing need for skills and qualifications. However, it has been shown that, at times, there are major mismatches in the labour market, which translate into a waste of resources and situations of great inefficiency, both from the perspective of individuals and from that of society as a whole. There is, therefore, a pressing need to enhance the coordination of education, economic and industrial policies. Although far from perfect, efforts at forecasting future demands for skills may be useful tools for reducing labour market mismatches and for guiding education policies (CEDEFOP, 2010 [36] and CEDEFOP, 2012 [37]). An alternative is to create a framework that can define the necessary level of competencies and abilities required at each level of formal education and to ensure that these coincide with labour market requirements.

To conclude, this thesis opens up the way for the development of a number of different research lines. One obvious extension is the analysis of other sources of labour market discrimination. Here, the gender inequality associated with occupational mismatch has been explored in-depth, but it is clearly of great interest to analyse other sources of discrimination, such as that linked to a worker's country of origin, with particular emphasis on the differences between the return to the productive characteristics of national and immigrant workers. A second extension of value would be the analysis of the incidence of gender inequality within specific countries and regions, exploring differences in the distribution of wage discrimination by the observable characteristics of workers. Finally, the publication of the second wave of PIAAC in 2021 will provide scholars with the unique opportunity to explore cohort dynamics in the labour market.

References

[1] Addabbo, T. and Favaro, D. (2011). Gender wage differentials by education in Italy. *Applied Economics*, 43:4589–4605.

[2] Aikens, N. and Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, 100(2):235–251.

[3] Albert, C. (2000). Higher education demand in Spain: The influence of labour market signals and family background. *Higher Education*, 40(2):147–162.

[4] Allen, J. and van der Velden, R. (2001). Educational mismatches versus skill mismatches: Effects on wage, job satisfaction, and on-the-job search. *Oxford Economic Papers*, 53:434–452.

[5] Allen, J., van der Velden, R., and Levels, M. (2013). *Skill mismatch and use in developed countries: Evidence from the PIAAC study*. Maastricht University, School of Business and Economics, Maastricht, Netherlands.

[6] Ananat, E., Gassman-Pines, A., Francis, D., and Gibson-Davis, C. (2011). Children left behind: The effects of statewide job loss on student achievement. *NBER Working Paper No. 17104*.

[7] Angelov, N., Johansson, P., and Lindahl, E. (2013). Is the persistent gender gap in income and wage due to unequal family responsibilities? *Discussion Paper No. 7181, Bonn, IZA*.

[8] Angrist, J. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

[9] Aparicio, A. (2010). High-school dropouts and transitory labor market shocks: The case of the spanish housing boom. *Discussion Paper No. 5139, Bonn, IZA*.

[10] Arranz, J. and García-Serrano, C. (2012). How green was my valley. *International Journal of Manpower*, 35(7):1059–1087.

[11] Arulampalam, W., Booth, A., and Bryan, M. (2007). Education and skill mismatches: wage and job satisfaction consequences. *International Journal of Manpower*, 34(5):416–428.

[12] Aughinbaugh, A. (2012). The effects of high school math curriculum on college attendance: Evidence from the nlsy97. *Economics of Education Review*, 31(6):861–870.

[13] Badillo-Amador, L. and Vila, L. (2013). Education and skill mismatches: Wage and job satisfaction consequences. *International Journal of Manpower*, 34(5):416–428.

[14] Barrington, B. and Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts and nongraduates. *Journal of Educational Research*, 86(6):309–319.

[15] Bauer, T. (2000). Educational mismatch and wages: A panel analysis. *Economics of Education Review*, 21:221–229.

[16] Becker, G. (1964). *Human Capital. 2nd edition*. Columbia University Press, New York.

[17] Belzil, C. and Leonardi, M. (2007). Can risk aversion explain schooling attainments? Evidence from Italy. *Labour Economics*, 14(6):957–970.

[18] Berger, L., Hill, J., and Waldfolgel, J. (2005). Maternity leave, early maternal employment and child health and development in the us. *The Economic Journal*, 115:F29–F47.

[19] Bertrand, M. and Pan, J. (2011). The trouble with boys: Social influences and the gender gap in disruptive behavior. *NBER Working Paper No. 17541*.

[20] Bickel, R. (1989). Opportunity and high school completion. *The Urban Review*, 21(4):61–84.

[21] Black, D., McKinnish, T., and Sanders, S. (2005). The economic impact of the coal boom and bust. *The Economic Journal*, 115(503):449–476.

[22] Blau, F. and Kahn, L. (2016). The gender wage gap: Extent, trends, and explanations. *NBER Working Paper No. 21913*.

[23] Bloom, D., Canning, D., Flink, G., and Finlay, J. (2009). Fertility, female labor force participation, and the demographic dividend. *Journal of Economic Growth*, 14(2):79–101.

[24] Brunello, G. and Rocco, L. (2006). The effect of migration on the school performance of natives: Cross country evidence using PISA test scores. *Economics of Education Review*, 32:234–246.

[25] Budria, S. and Moro-Egido, A. (2008). Education, educational mismatch, and wage inequality: Evidence for Spain. *Economics of Education Review*, 27:332–341.

[26] Budria, S. and Moro-Egido, A. (2014). Overqualification, skill mismatches and wages in private sector employment in Europe. *Technological and Economic Development of Economy*, 20(3):457–483.

[27] Calero, J., Choi, A., and Waisgrais, S. (2010). Determinantes del riesgo de fracaso escolar en España: Una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006. *Revista de Educación*, pages 225–256.

[28] Calero, J. and Escardíbul, J. (2007). Evaluación de Servicios Educativos: El rendimiento en los centros públicos y privados medido en PISA-2003. *Hacienda Pública Española*, 183:33–66.

[29] Calero, J and Waisgrais, S (2009). *Rendimientos Educativos de los alumnos inmigrantes: identificación de la incidencia de la condición de inmigrante y de los peer effect*. XVI Encuentro de Economía Pública, 2009-01-01, ISBN 978-84-691-8950-4.

[30] Calsamiglia, C. and Güell, M. (2014). The illusion of school choice: Empirical evidence form Barcelona. *Discussion Paper No. 8202, Bonn, IZA*.

[31] Camarata, S. and Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, 34:231–252.

[32] Card, D. and Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *The Quarterly Journal of Economics*, 116(2):705–746.

[33] Carneiro, P. and Heckman, J. (2002). The evidence on credit constraints in post-secondary schooling. *The Economic Journal*, 112(482):705–734.

[34] Carrington, B., Tymss, P., and Merrell, C. (2008). Role models, school improvement and the gender gap.- do men bring out the best in boys and women the best in girls?,. *British Educational Research Journal*, 34(3):315–327.

[35] Cebrián, I. and Moreno, G. (2013). Labour market intermittency and its effect on gender wage gap in Spain. *Review d'Interventions économiques*, 47:2–19.

[36] CEDEFOP (2010). *Skills supply and demand in Europe: medium-term forecast up to 2020*. European Centre for the Development of Vocational Training. Publications Office of the European Union, Luxembourg.

[37] CEDEFOP (2012). *Future skills supply and demand in Europe*. European Centre for the Development of Vocational Training. Publications Office of the European Union, Luxembourg.

[38] Checchi, D. and van de Werfhorst, H. (2014). Educational policies and income inequality. *Discussion Paper No. 8222, Bonn, IZA*.

[39] Chevalier, A. (2003). Measuring mismatch. *Economica*, 279(70):509–531.

[40] Chingos, M. (2010). The impact of a Universal Class-Size Reduction Policy: Evidence from Florida's Statewide Mandate. *Program on Education Policy and Governance. Working Papers Series. PEPG 10-03. Harvard University.*

[41] Choi, A. and Calero, J. (2011). Determinantes del riesgo de fracaso escolar en España en PISA-2009 y propuestas de reforma. *Revista de Educación*, (362):562–593.

[42] Claxton, G. and Lucas, B. (2009). *School as a foundation for lifelong learning: the implications of a lifelong learning perspective for the re-imagining of school-age education*. Inquiry into the Future of Lifelong Learning (IFLL) sector paper 1, National Institute of Adult and Continuing Education, Leicester, UK.

[43] Cohen, P. (1987). *A Critical Analysis and Reanalysis of the Multisection Validity meta-analysis*. Annual meeting of the American Educational Research Association 1987.

[44] Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., and York, R. (1966). *Equality of Education Opportunity*. U.S. Government Printing Office, Washington D.C.

[45] Cordero, J., Crespo, E., and Pedraja, F. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España. *Revista de Educación*, 362.

[46] Cordero, J., Manchón, C., and García, M. (2011). Los resultados educativos españoles en PISA 2009 y sus condicionantes. *Investigaciones de Economía de la Educación*, 6:70–87.

[47] Crawford, C., Dearden, L., and Greaves, E. (2011). *Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills in England*. Institute for Fiscal Studies - Nuffield Foundation.

[48] Defloor, B., Van Ootegem, L., and Verhofstadt, E. (2015). A good or bad transition from school to work: Who is responsible? *International Journal of Manpower*, 31(8):1207–1226.

[49] Desjardins, R. and Rubenson, K. (2011). An analysis of skill mismatch using direct measures of skills. *OECD Education Working Papers. Paris: Organisation for Economic Co-operation and Development*.

[50] Dolton, P. and Vignoles, A. (2000). The incidence and effects of overeducation on the U.K. graduate labor market. *Economics of Education Review*, 19:179–198.

[51] Duryea, S. and Arends-Kuenning, M. (2003). School attendance, child labor, and local labor markets in urban Brazil. *World Development*, 31(7):1165–1178.

[52] Ehrenberg, R. and Brewer, D. (1994). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, 14(1):1–21.

[53] Entorf, H. and Minoiu, N. (2005). What a difference immigration policy makes: A comparison of PISA scores in europe and traditional countries of immigration. *German Economic Review*, 6(3):355–376.

[54] Entwisle, D., Alexander, K., and Olson, L. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology*, 110(5):1458–1502.

[55] Falch, T., Nyhus, O., and Strøm, B. (2014). Causal effects of mathematics. *Labour Economics*, 31:174–187.

[56] Favaro, D. and Magrini, S. (2008). Group versus individual discrimination among young workers: a distributional approach. *Journal of Socio-Economics*, 37:1856–1879.

[57] Ferguson, R. and Ladd, H. (1996). *Holding Schools Accountable. How and Why Money Matters: An Analysis of Alabama Schools*. Helen Ladd, ed., Brookings institution.

[58] Fernández, E., Mena, L., and Riviere, J. (2010). School failure and dropouts in Spain. *Social Studies Collection, Obra Social Fundación La Caixa*, (29).

[59] Flisi, S., Goglio, V., Meroni, E., Rodrigues, M., and Vera-Toscano, E. (2013). Measuring occupational mismatch: Overeducation and overskill in Europe – Evidence from PIAAC. *Social Indicators Research*, pages 1–39.

[60] Flisi, S., Goglio, V., Meroni, E., Rodrigues, M., and Vera-Toscano, E. (2014). *Occupational mismatch in Europe: Understanding overeducation and overskillin for policy making*. Joint Research Centre – European Commission, Luxembourg.

[61] Fortin, N. (2005). Gender role attitudes and the labour-market outcomes of women across OECD countriess. *Oxford Review of Economic Policy*, 21(3):416–438.

[62] Fryer, R. (2016). The 'pupil' factory: Specialization and the production of Human Capital in schools. *NBER Working Paper No. 22205*.

[63] García, M. (2011). *Nivel Socioeconómico, tipo de Escuela y Resultados Educativos en España: El Caso de TIMSS PIRLS 2011. Estudio Internacional de Progreso en Comprensión Lectora, Matemáticas y Ciencias*, volume 2. Center for Research on Child Wellbeing. Princeton University.

[64] Gautier, P. (2002). Unemployment and search externalities in a model with heterogeneous jobs and workers. *Economica*, 69(273):21–40.

[65] Green, F. and McIntosh, S. (2007). Is there a genuine under-utilization of skills amongst the over-qualified? *Applied Economics*, 39(4):427–439.

[66] Groot, W. (1996). The incidence of, and returns to overqualification on the UK. *Applied Economics*, 28(10):1345–1350.

[67] Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19:F64–F98.

[68] Hanushek, E. (2003). The failure of input-based schooling policies. *The Economic Journal*, 13:F64–F98.

[69] Haveman, R. and Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature, American Economic Association*, 33(4):1829–1878.

[70] Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

[71] Holmes, C. (1989). *Grade level retention effects: A meta-analysis of research studies - Flunking grades: Research and policies on retention*. L. A. Shepard M. L. Smith (Eds.), London:Falmer.

[72] INE (2015). *Labor Force Survey (LFS) Education statistics. Level of education, training and permanent abandonment: Exploration of educational variables of labour force survey"*. Instituto Nacional de Estadística. Available from: https://www.educacion.gob.es/educabase/menu.do?type=pcaxispath=/Formacionyml/EPA2014/Indifile=pcaxisl=s0.

[73] Iriondo, I. and Pérez-Amaral, T. (2016). The effect of educational mismatch on wages in Europe. *Journal of Policy Modeling*, 38(2):304–323.

[74] Jimmerson, S. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3):420–437.

[75] Joensen, J. and Nielsen, H. (2009). Is there a causal effect of high school math on labor market outcomes. *Journal of Human Resources*, 44(1):171–198.

[76] Knight, J. and Sabot, R. (1987). The rate of return on educational expansion. *Economics of Education Review*, 6(3):255–262.

[77] Krueger, A. (2002). Economic considerations and class size. *The Economic Journal*, 113(485):F34–F63.

[78] Lamo, A. and Messina, J. (2010). Formal education, mismatch and wages after transition: Assessing the impact of unobserved heterogeneity using matching estimators. *Economics of Education Review*, 29:1086–1099.

[79] Lassibille, G. and Navarro Gómez, L. (2008). Why do higher education students drop out? Evidence from Spain. *Education Economics*, 16(1):89–105.

[80] Lavy, V. (2012). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *NBER Working Paper No. 16227*.

[81] Lee, V. and Burkam, D. (2002). *Inequality at the Starting Gate*. Economic Policy Institute, Washington, DC.

[82] Li, M., Fox, J., and Grieshaber, S. (2017). *Contemporary Issues and Challenge in Early Childhood Education in the Asia-Pacific Region*. New Frontiers of Educational Research Series. Springer, Singapore.

[83] Liu, K., Salvanes, K., and Sørensen, E. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84(C):3–17.

[84] Maani, S. and Kalb, G. (2007). Academic performance, childhood economic resources, and the choice to leave school at age 16. *Economics of Education Review*, 26(3):361–374.

[85] Machado, J. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20:445–465.

[86] Malley, J. and Muscatelli, V. (1999). Business cycles and productivity growth: Are temporary downturns productive or wasteful? *Research in Economics*, 53(4):337–364.

[87] Manski, C. and Wise, D. (1983). *College Choice in America*. Harvard University Press, Cambridge, Massachusetts.

[88] Martin, A. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology*, 101(1):101–114.

[89] Mavromaras, K., McGuiness, S., Richardson, S., Sloane, P., and Weiz, Z. (2011). *Over-skilling and job satisfaction in the Australian labor force.* National Centre for Vocational Education Research, Adelaide.

[90] McCarthy, K. (1998). *Adaptation of Immigrant Children to the United States: A Review of the literature*. Center for Research on Child Wellbeing. Princeton University, USA.

[91] McGuiness, S. (2003b). Graduate overqualification as a sheepskin effect: Evidence from Northern Ireland. *Applied Economics*, 35(5):597–608.

[92] McGuiness, S. and Sloane, P. (2011). Labour market mismatch among UK graduates: An analysis using REFLEX data. *Economics of Education Review*, 30:130–145.

[93] McIntosh, S. and Vignoles, A. (2001). Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers*, 53(3):453–481.

[94] McNeal, R. (1997). High school dropouts: A closer examination of school effects. *Social Science Quarterly*, 78(1):210–222.

[95] MECD (2009). *Panorama de la educación. Indicadores de la OCDE 2009*. Ministerio de Educación, Cultura y Deporte, Madrid, Spain.

[96] Micklewright, J., Pearson, M., and Smith, S. (1990). Unemployment and early school leaving. *The Economic Journal*, 100(400):163–169.

[97] Mincer, J. (1974). *Schooling, experience and earnings*. National Bureau of Economic Research, New York.

[98] Montmarquette, C., Viennot-Briot, N., and Dagenais, M. (2007). Dropout, school performance, and working while in school. *The Review of Economics and Statistics*, 89(4):752–760.

[99] Nordin, M., Persson, I., and Dan-Olof, R. (2010). Education-occupation mismatch: Is there an income penalty? *Economics of Education Review*, 29(6):1047–1059.

[100] Oaxaca, R. (1979). Male-female wage differentials in urban labour markets. *International Economics Review*, 14:693–709.

[101] OECD (2000). *PISA 2000: Technical Report*. OECD, Paris.

[102] OECD (2003). *PISA 2003: Technical Report PISA 2003*. OECD, Paris.

[103] OECD (2006). *PISA 2006: Technical Report*. OECD, Paris.

[104] OECD (2009). *Resultados del Informe PISA 2009 - Aprendiendo a Aprender (Volumen V)*. OECD, Paris.

[105] OECD (2009a). *PISA Data Analysis Manual. SPSS Second Edition*. OECD, Paris.

[106] OECD (2009b). *PISA 2009: Technical Report PISA 2009*. OECD, Paris.

[107] OECD (2010). *Education at Glance*. OECD, Paris.

[108] OECD (2011). *Education at Glance*. OECD, Paris.

[109] OECD (2013a). *PISA 2009: Technical Report of the Survey of Adult Skills (PIAAC)*. OECD, Paris.

[110] OECD (2013b). *OECD Skills outlook 2013: First results from the survey of adult skills*. OECD, Paris.

[111] OECD (2014). *PISA 2012 Technical Report*. OECD, Paris.

[112] Patrinos, H. (2016). Estimating the return to schooling using the mincer equation. *IZA World of Labor, Institute for the Study of Labor*.

[113] Peraita, C. and Pastor, M. (2000). The primary school dropout in Spain: the influence of family background and labor market conditions. *Education Economics*, 8(2):157–168.

[114] Perry, A., Wiederhold, S., and Ackermann-Piek, D. (2014). How can skill mismatch be measured? new approaches with PIAAC. *Methods, data, analyses*, 8(2):137–174.

[115] Petrongolo, B. and San Segundo, M. (2002). Staying-on at school at 16: the impact of labor market conditions in Spain. *Economics of Education Review*, 21(4):353–365.

[116] Pfeffermann, D., Skinner, C., Holmes, D., Goldstein, H., and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *JRSS, Series B*, 60:123–140.

[117] Quinn, M. and Rub, S. (2006). Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries. *Economics of Education Review*, 25:147–156.

[118] Quintini, G. (2011a). Over-qualified or under-skilled: A review of existing literature. *OECD Social, Employment and Migration Working Papers. Paris: OECD*.

[119] Raaum, O. and Røed, K. (2006). Do business cycle conditions at the time of labor market entry affect future employment prospects? *The Review of Economics and Statistics*, 88(2):193–210.

[120] Rakowska, A. and Cichorzewska, M. (2016). *Competences needed on the future labour market-Results from Delphi method*. Paper presented at Management, Knowledge and Learning – Technology, Innovation and Industrial Management, Joint International Conference 2016. doi: http://www.toknowpress.net/ISBN/978-961-6914-16-1/papers/ML16-167.pdf, Timisora, Romania.

[121] Raudenbusch, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Second Edition. Sage Publications Inc.

[122] Rees, D. and Mocan, H. (1997). Labor market conditions and the high school dropout rate: Evidence from New York State. *Economics of Education Review*, 16(2):103–109.

[123] Reynolds, A. and Temple, J. (1998). Extended early childhood intervention and school achievement: Age 13 findings from the Chicago Longitudinal Study. *Child Development*, 69:231–246.

[124] Rivkin, M. (1995). *The Great Outdoors: Restoring Children's Right to Play Outside*. NAEYC, Washington.

[125] Robertson, E. (2011). The effects of quarter of birth on academic outcomes at the elementary school level. *Economics of Education Review*, 30:300–311.

[126] Robst, J. (2007). Education and job match: The relatedness of college major and work. *Economics of Education Review*, 26(4):397–407.

[127] Rubin, D. (1987). *Multiple Imputation for non-response in surveys*. John Wiley Sons, New York.

[128] Rumberger, R. (2011). *Dropping Out*. Harvard University Press, Cambridge, Massachusetts.

[129] Salinas-Jiménez, M., Rahona-López, M., and Murillo-Huertas, I. (2013). Gender wage differentials and educational mismatch: An application to the spanish case. *Applied Economics*, 45(30):4226–4235.

[130] Santos, M. and Sequeira, T. (2014). Skills mismatch and wage inequality: Evidence for different countries in Europe. *Technological and Economic Development of Economy*, 19:425–453.

[131] Sattinger, M. (1993). Generating inequality. *Journal of Economic Literature*, 31:831–880.

[132] Schady, N. (2004). Do macroeconomic crises always slow human capital accumulation? *World Bank Economic Review*, 18(2):131–153.

[133] Sleicher, A. (2007). Can competencies assessed by pisa be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, 8(4):349–357.

[134] Slonimczyk, F. (2013). Earnings inequality and skill mismatch in the U.S.: 1973–2002. *The Journal of Economic Inequality*, 11(2):163–194.

[135] Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87:355–374.

[136] Sprietsma, M. (2010). Effect of relative age in the first grade of primary school on long-term scholastic results: international comparative evidence using PISA 2003. *Education Economics*, 18(1):1–32.

[137] Stern, D., Paick, I., Catterall, J., and Nakata, Y. (1989). Labor market experience of teenagers with and without high school diplomas. *Economics of Education Review*, 8(3):233–246.

[138] Strauss, J. and Thomas, D. (1995). *Human resources: Empirical modeling of household and family decisions, note = Handbook of Development Economics, Volume IIIA*. edited by J. Behrman and T.N. Srinivasan, Amsterdam: Elsevier.

[139] Thurow, L. (1975). *Generating inequality*. Basic Books, New York.

[140] Verdugo, R. and Verdugo, N. (1989). The impact of surplus schooling on earnings: Some additional findings. *The Journal of Human Resources*, 24(4):629.

[141] Warren, J. and Lee, J. (2003). The impact of adolescent employment on high school dropout: differences by individual and labor - market characteristics. *Social Science Research*, 32(1):98–128.

[142] Zhou, A., Ma, X., and Aoyagi, H. (2007). Parental expectation of early childhood education: comparison between China, Japan, and Korea. *Frontiers of Education in China*, 2(1):140–147.

Appendix A

Annexes Chapter 2

Table A.1 Variance reduction analysis results

2000	Variance component	s.d.	Chi-square	P-value
Null model	0.845	0.919	871.38	0.000
Full model	0.326	0.569	413.72	0.000
2003	Variance component	s.d.	Chi-square	P-value
Null model	0.855	0.925	1,479.38	0.000
Full model	0.478	0.690	1,005.15	0.000
2006	Variance component	s.d.	Chi-square	P-value
Null model	0.838	0.916	2,716.51	0.000
Full model	0.465	0.681	2,055.64	0.000
2009	Variance component	s.d.	Chi-square	P-value
Null model	1.000	1.000	4,697.29	0.000
Full model	0.583	0.763	3,246.37	0.000

Note: The table reports the difference between the variance component for the unconditional model with random intercept (one-way ANOVA) and the full model. The table reports information about the outcome variability of within-group and between-group variance. The significant difference in the variance between groups in the four estimations justifies the use of hierarchical models.

Appendix B

Annexes Chapter 3

Table B.1 Variable description

Variable	Description
Gender	Dummy variable identifying females
Household unemployment	Dummy variable identifying if both parents are unemployed
First generation immigrants	Dummy variable identifying if student and parents are born outside of Spain
Second generation immigrants	Dummy variable identifying if student was born in Spain but parents are born outside of Spain
ESCS index	Index of economic, social and cultural status
Unemployment rate of the fathers in the school	Ratio of unemployed fathers / total number of fathers in the school
Unemployment rate of the mothers in the school	Ratio of unemployed mothers / total number of mothers in the school
Average ESCS by school	Average of ESCS index by school
Community (<100.000 inh.)	Dummy variable identifying if the community where the school is located has less than 100,000 inhabitants
Community (>1.000.000 inh.)	Identifies if the community where the school is located has more than 1,000,000 inhabitants
School private public dependent	Private schools financed with public resources
School public	Public schools
Percentage of females by school	Percentage of females over the school population
Immigrant ratio (over 30%)	Schools were the concentration of immigrants is over 30% of the school population
Student/teacher ratio	Number of students divided by the number of teachers on a given grade by school
Autonomy on budget allocation	If the school has the authority and responsibility to decide the management allocation of the funds
Autonomy of course content	If the school has the authority and responsibility to decide the curriculum and instructional assessment
Household unemployment*4th ESCS quartile	Interaction between household unemployment and the 4th quartiles of the ESCS index.
Household unemployment*3rd ESCS quartile	Interaction between household unemployment and the 3rd quartiles of the ESCS index.
Household unemployment*2nd ESCS quartile	Interaction between household unemployment and the 2nd quartiles of the ESCS index.

Source: Own elaboration

Table B.2 Variance reduction analysis results for the models displayed in Table 3.2

	2006 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.201	0.865	6,201.95	0.00			
Full model	0.073	0.673	2,635.96	0.00	0.073	0.710	0.202
	2009 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.208	0.869	9,441.14	0.00			
Full model	0.118	0.774	4,607.03	0.00	0.100	0.542	0.192
	2012 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.207	0.861	7,547.00	0.00			
Full model	0.083	0.692	3,328.00	0.00	0.097	0.687	0.219
	Pooled model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.209	0.868	23,456.97	0.00			
Full model	0.097	0.732	11,138.21	0.00	0.091	0.632	0.204

Note: The table reports the difference between the variance component for the unconditional model with random intercept (one-way ANOVA) and the full model. It also reports information about the outcome variability of within-group and between-group variance. The significant difference in the variance between groups in all the estimations justifies the use of linear hierarchical models.

Table B.3 Variance reduction analysis results for the models displayed in Table 3.4

	2006 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.107	0.459	1,231.42	0.00			
Full model	0.047	0.273	884.61	0.00	0.084	0.628	0.142
	2009 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.143	0.512	2,486.10	0.00			
Full model	0.101	0.427	1,751.94	0.00	0.081	0.377	0.123
	2012 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.133	0.490	2,099.85	0.00			
Full model	0.069	0.333	1,491.82	0.00	0.046	0.543	0.113
	Pooled model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.129	0.490	5,720.09	0.00			
Full model	0.081	0.378	4,259.82	0.00	0.067	0.442	0.115

Note: The table reports the difference between the variance component for the unconditional model with random intercept (one-way ANOVA) and the full model. It also reports information about the outcome variability of within-group and between-group variance. The significant difference in the variance between groups in all the estimations justifies the use of linear hierarchical models.

Table B.4 Variance reduction analysis results for the models displayed in Table 3.5

	2006 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.171	0.503	1,979.39	0.00			
Full model	0.101	0.378	1,349.43	0.00	0.015	0.460	0.091
	2009 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.215	0.560	2,832.27	0.00			
Full model	0.167	0.495	2,062.03	0.00	0.021	0.281	0.077
	2012 model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.184	0.524	2,356.74	0.00			
Full model	0.167	0.495	1,839.40	0.00	0.030	0.431	0.104
	Pooled model						
	Intra class correlation	Reliability range	Chi-square	P-value	Within school variance	Between school variance	Total explained variance
Null model	0.188	0.526	7,058.26	0.00			
Full model	0.135	0.442	5,360.78	0.00	0.020	0.343	0.081

Note: The table reports the difference between the variance component for the unconditional model with random intercept (one-way ANOVA) and the full model. It also reports information about the outcome variability of within-group and between-group variance. The significant difference in the variance between groups in all the estimations justifies the use of linear hierarchical models.

Appendix C

Annexes Chapter 4

Table C.1 Descriptive statistics by country

Country	Male (%)	Female (%)	Age	S.D.	Years of formal education	S.D.	Logarithm of hourly wage	S.D.
Belgium	49.42	50.58	40.71	14.2	12.28	2.9	2.76	0.4
Chile	42.17	57.83	39.10	14.2	11.32	3.3	7.87	0.8
Czech Republic	45.38	54.62	38.65	15.6	13.02	2.8	4.67	0.6
Denmark	49.33	50.67	43.99	14.7	12.78	2.8	5.14	0.5
France	49.05	50.95	41.84	14.3	11.38	3.6	2.50	0.5
Greece	45.08	54.92	40.92	13.1	12.13	3.3	1.81	0.5
Indonesia	36.67	63.33	38.26	12.6	10.73	3.4	9.31	0.9
Israel	50.38	49.62	37.13	14.3	12.29	3.0	3.69	0.9
Italy	48.37	51.63	42.26	13.2	11.51	3.9	2.37	0.5
Japan	47.69	52.31	41.81	14.1	13.07	2.4	7.21	0.8
Lithuania	39.92	60.08	43.00	14.1	13.38	2.7	2.32	0.7
Netherlands	49.24	50.76	41.78	14.4	13.13	2.7	2.71	0.7
New Zealand	43.18	56.82	39.73	14.4	13.58	2.5	3.12	0.5
Norway	51.77	48.23	39.86	14.2	14.14	2.6	5.30	0.5
Poland	50.53	49.47	31.23	13.8	12.47	2.7	2.56	0.6
Republic of Korea	46.53	53.47	40.57	13.7	12.61	3.3	9.28	0.7
Slovakia	47.28	52.72	39.42	14.3	12.76	2.8	1.32	0.7
Slovenia	49.07	50.93	41.46	14.1	10.43	2.0	1.95	0.4
Spain	48.95	51.05	40.08	13.6	11.15	3.5	2.17	0.6
United Kingdom	42.03	57.97	41.20	13.6	13.18	2.3	2.36	0.6

Note: Based on PIAAC database

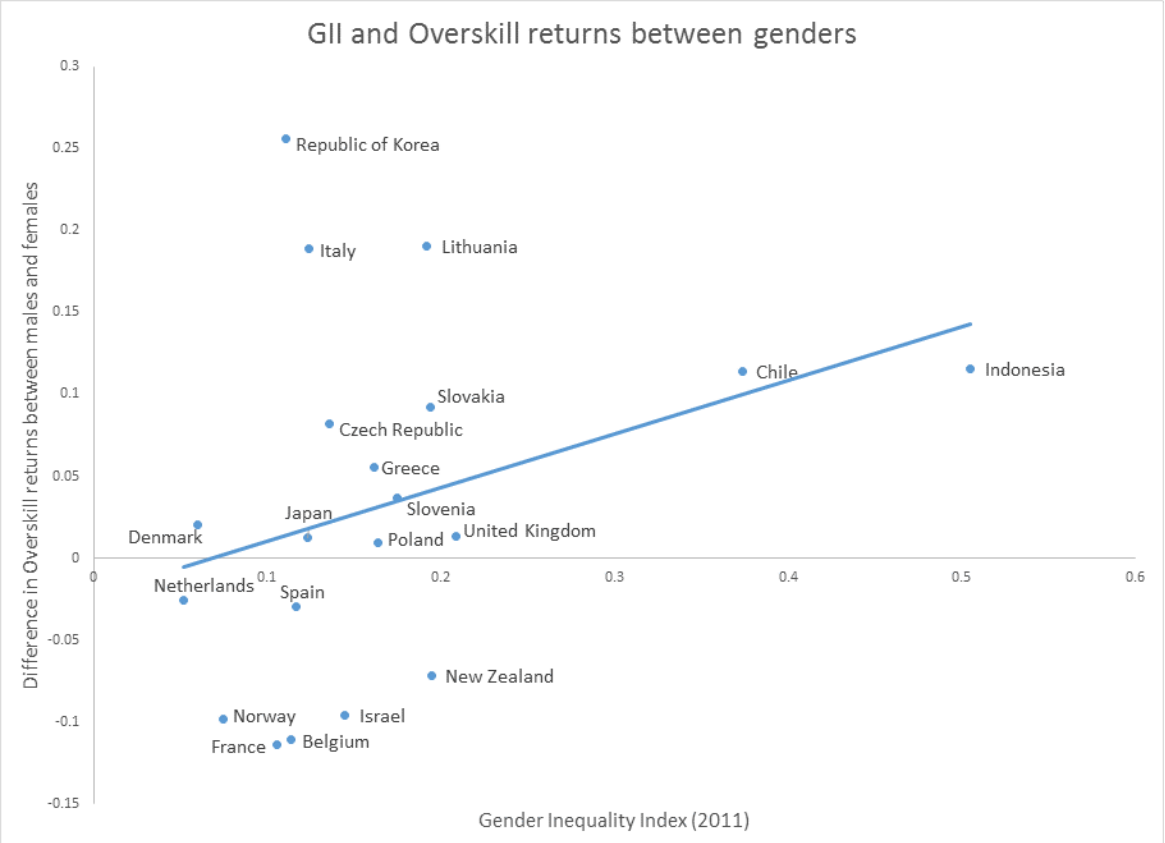

Fig. C.1 Difference in returns to overqualification and Gender Inequality Index
Source: Own calculations and UNDP data (2011).

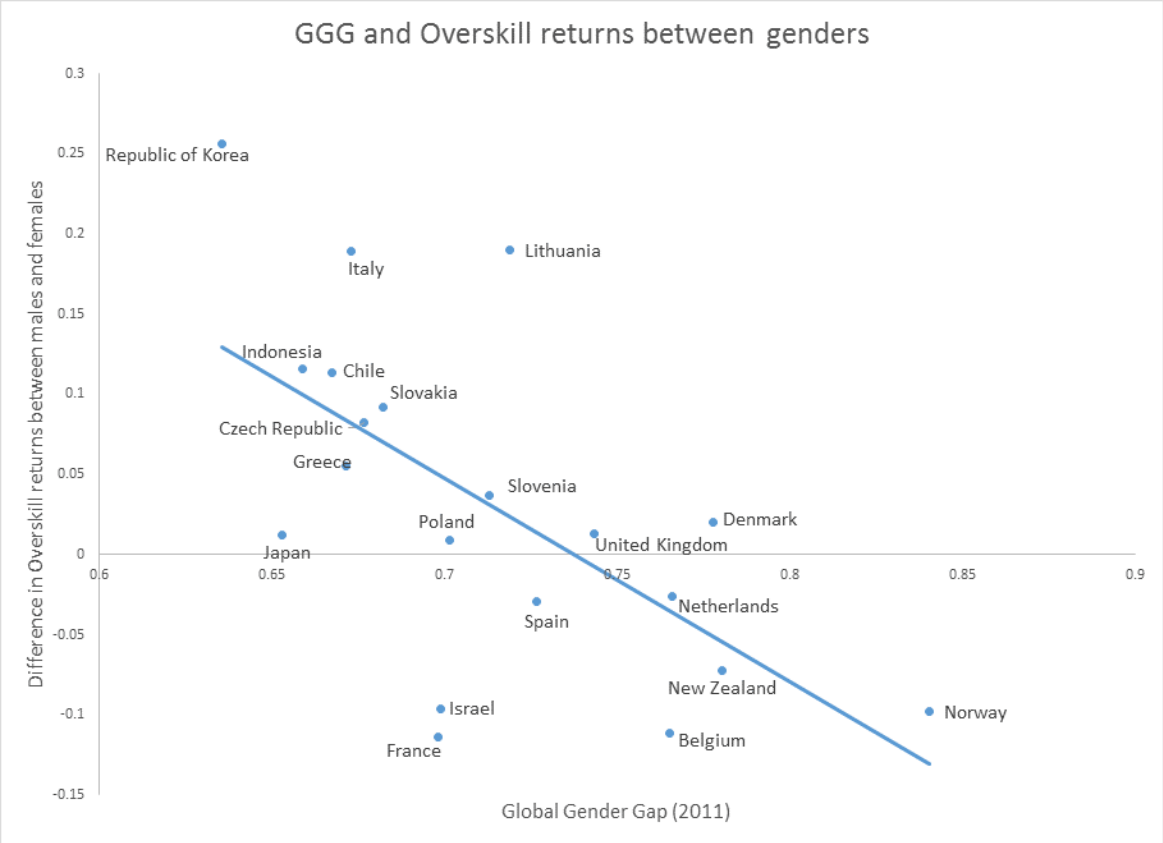

Fig. C.2 Difference in returns to overqualification and Global Gender Gap Index
Source: Own calculations and World Economic Forum data (2011)

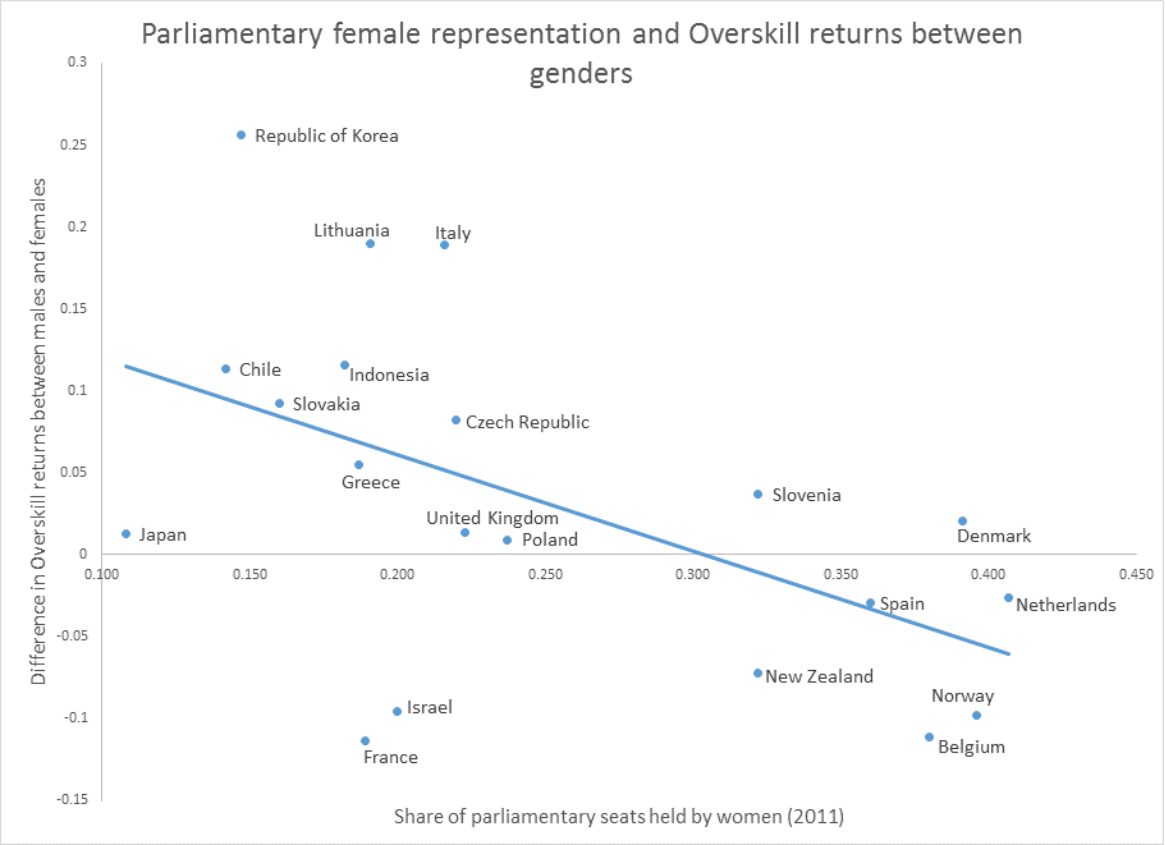

Fig. C.3 Difference in returns to overqualification and parliamentary female representation
Source: Own calculations and World Bank data (2011)

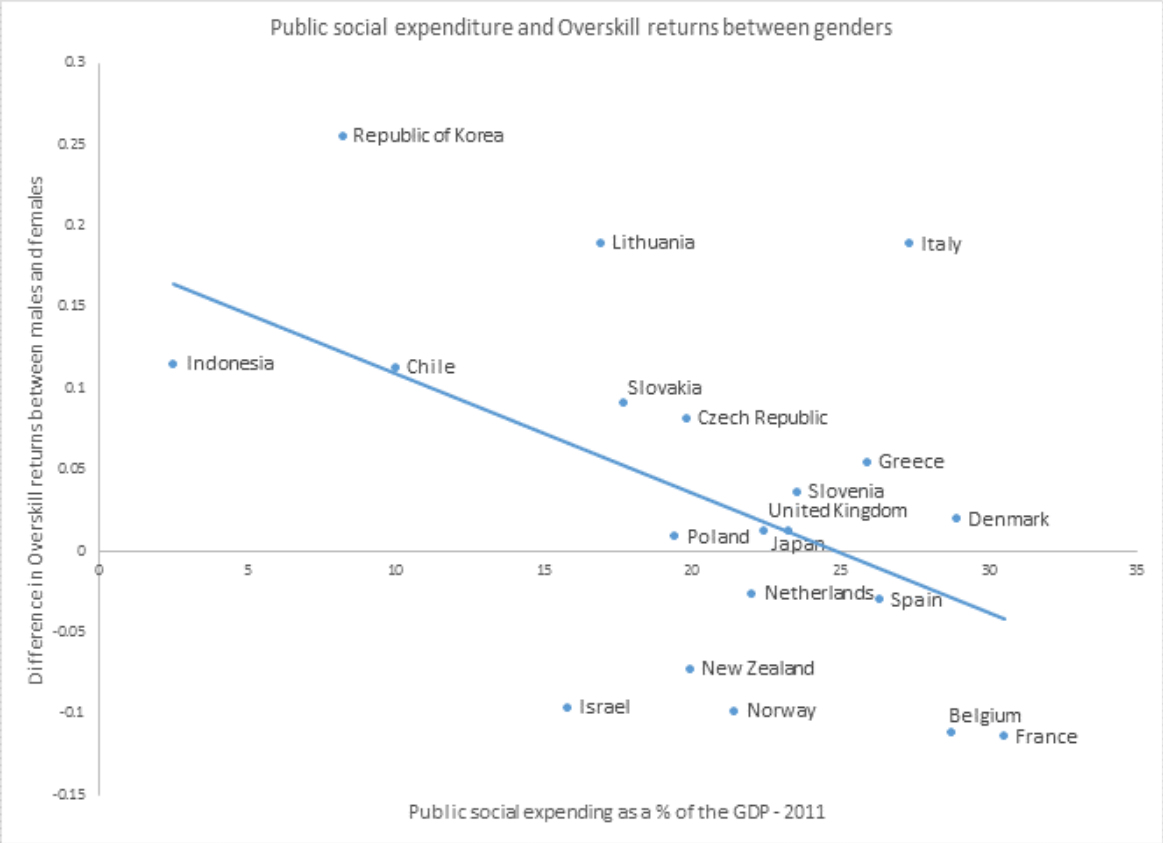

Fig. C.4 Difference in returns to overqualification and public social expenditure
Source: Own calculations and OECD data (2011)

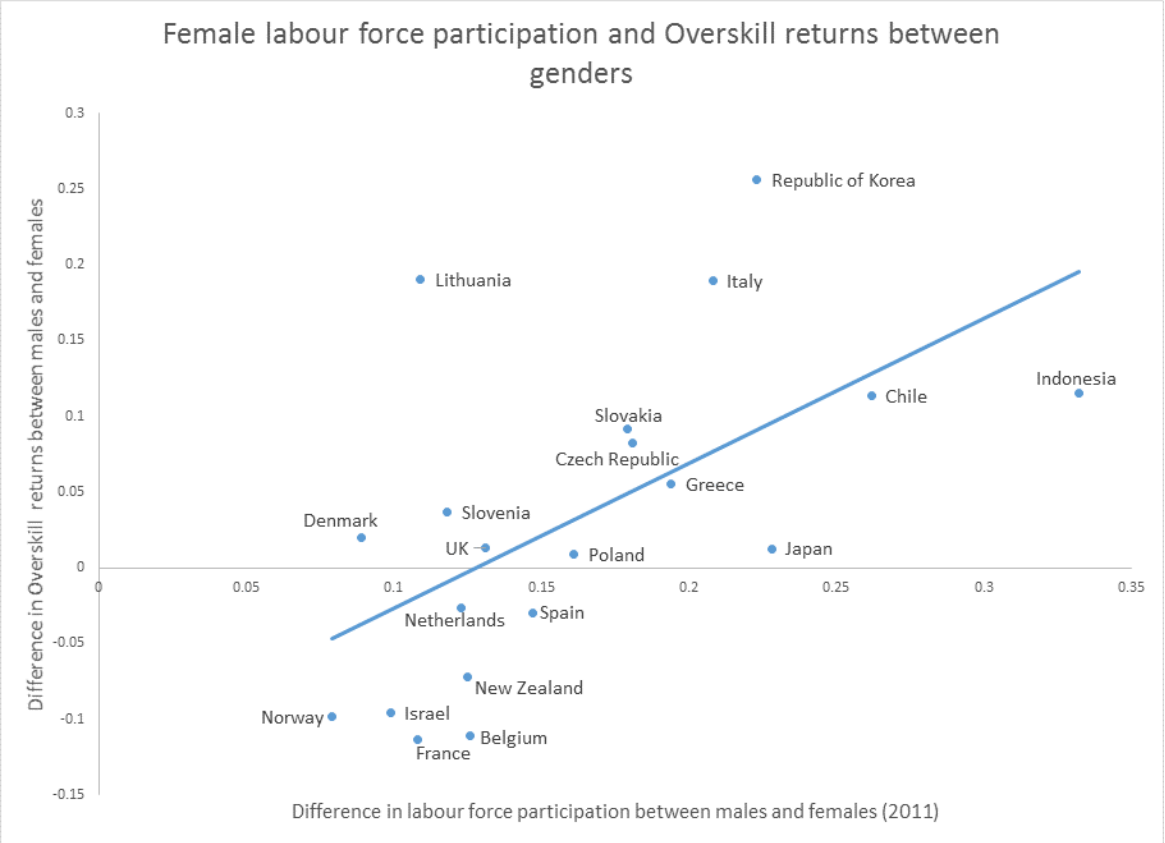

Fig. C.5 Difference in returns to overqualification and female labour force participation
Source: Own calculations and World Bank data (2011)