Essays on Overeducation: Evidence from Spain

Sandra Nieto Viramontes

PhD in Economics

Essays on Overeducation: Evidence from Spain

Sandra Nieto Viramontes

PhD in Economics | Sandra Nieto Viramontes

Universitat de Barcelona

PhD in Economics

Thesis title:

Essays on Overeducation: Evidence from Spain

PhD student:
Sandra Nieto Viramontes

Advisor:
Raúl Ramos Lobo

Date:
July 2014

Universitat de Barcelona

A mis padres y a mis hermanos

Acknowledgements

First and foremost, I would like to express my most heartfelt gratitude to my supervisor Raúl Ramos. I really appreciate his guidance, support and dedication all the time. I owe this thesis and everything that derives from it to him. Moltes gràcies Raul!

I would like to thank the Department of Econometrics, Statistics and Spanish Economy for its logistic support and the Catalan and the Spanish Government for the financial support. Additionally, I would like to thank Jordi Roca, from UB-Economics, for his patience and help. I want to express my gratitude to the AQR Research Group and its director Enrique López-Bazo for the support and encouragement during the last months of writing the thesis. I am also grateful to Jordi Suriñach for introducing me into the AQR and showing me the existence of research.

I would also like thank Dr Joop Hartog for his valuable comments during my stay at the University of Amsterdam. I am also very grateful to Dr Hessel Oosterbeek. It has been a pleasure for me to meet all of you.

During all this time, I have had the opportunity to know some researchers and, above all, wonderful people. In particular, I am grateful to Elisabet Motellón, Luis Díaz and Inés Murillo for their generosity, confidence and support.

Furthermore, I would like to thank all my colleagues at the University, with whom I have shared a lot of experiences. In particular, I thank Ernest Miguélez for his great company and the time we spent at the office. It was always funny and enjoyable.

I would also like to mention that I am very thankful to Esther Goya, Aïda Solé and Elena Costas for their friendship, support and the advice I received from them. I would also like to thank all my friends for their understanding that I could not always be around as I had to work on my thesis.

Finally, I am extremely grateful to my family, especially to my parents, who have sacrificed their lives to create so many opportunities for their children. They taught me to work hard, to be strong and grateful and never lose the smile.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Background and motivation

Investment in human capital is a key tool for the social and economic progress in every country. Thus one of the most important public policies in the last century has focused on increasing the level and quality of education among the population. This goal has been reached by all developed countries.

However, the rapid increase of the percentage of the population with high education in developed economies during the last decades has probably contributed to labour market inefficiency. It is observed that part of this population is not working in jobs that require their level of education; otherwise they end up in jobs that require workers with a lower level of education. This situation is called overeducation.

The concept of overeducation was introduced by Richard Freeman in his book *The Overeducated American*, which was published in 1976 (Freeman, 1976). From a macroeconomic point of view, he argued that the apparent declination of the income returns from high education since 1970 in the U.S. could be explained by an excess of supply of high-educated workers. Years later, Duncan and Hoffman (1981) made an important contribution analysing overeducation at an individual level by comparing workers whose level of attained education matches with that required by their jobs with workers who have more (and less) attained education. After that, the analysis was extended for several developed but also developing countries.

Nowadays, there are several contributions that focus on analysing the incidence of overeducation, its determinants and its consequences for workers who are suffering from it. Quantifying the incidence of overeducation is indeed crucial to understanding the importance of its negative consequences and to design public policies to reduce the overeducation weight on countries.

This thesis focuses on Spain, because it is a developed country supporting one of the largest percentages of overeducated workers (OECD, 2013), a feature that was also observed before the current economic crisis (OECD, 2009; and Verhaest and van der Velden, 2013). Spain is a country that was characterised by having a population with lower educational levels among the OECD countries in the 1960s and 1970s (Mas et al., 1995), and it has increased the percentage of its population with high education to the average of the OECD countries over three decades. However, the productive structure of the country is not comprised of mostly high-technology companies. This situation may point out an imbalance between the supply of and demand for high-skilled workers in the Spanish labour market. However, it has recently born a debate related to the skills and ability of workers in the country given that Spain is always at the bottom of the rankings that measure children's knowledge (PISA data) and adults' skills (PIAAC data). In this sense, overeducation may be a consequence of a lack of skills of workers, and policy implications should be targeted towards improving higher education institutions, i.e., universities.

The rest of this chapter is divided into four sections. Section 1.2 describes the different existing methods to measure overeducation. Section 1.3 summarises the most relevant findings of the consequences of overeducation on workers. Then, section 1.4 includes a short overview of the most supported theories to explain overeducation phenomena. Lastly, section 1.5 presents a summary of the three empirical analyses of the dissertation.

1.2. Measurement of educational mismatch

The most common way to approximate the concept of educational mismatch involves comparing the level of a worker's attained education with the level required by the worker's job. From this perspective, workers are overeducated (undereducated) if their level of education is higher (lower) than the level of education required by their jobs.

Literature on educational mismatch has proposed (at least) three different methods for measuring years of required education (see Hartog, 2000; and McGuinness, 2006 for a review): the objective or job analysis, the subjective or worker self-assessment (both direct and indirect) and the statistical or realised matches (in terms of the mean and the mode). In particular, the objective method is based on the opinion of expert analysts who determine what level of education workers should have in order to perform a certain job. A person is then overeducated (undereducated) if his/her level of education is higher (lower) than the level the analysts determine to be ideal for the occupation.

The subjective or self-assessment method relies on questions that ask workers about the schooling requirements of their job. Educational mismatch is determined by comparing workers' answers about required education and attained education. Workers are properly or well-matched when their attained education matches with their jobs' required education. Conversely, overeducated (undereducated) workers have more (less) attained education than required by their jobs.

Lastly, the statistical method based on the mean (Verdugo and Verdugo, 1989) considers that a person is overeducated (undereducated) if he/she has a level of education that is higher (lower) by more than one standard deviation than the mean level of education of the workers in that occupation. Nevertheless, Kiker et al. (1997) propose the use of the mode instead of the mean; so they consider a person who has a higher (lower) level of education than the mode for the job they perform to be overeducated (undereducated).

Unfortunately, at present there is no consensus as to which is the best method, because each has its advantages and disadvantages (Hartog, 2000). As a consequence, the use of one method or another usually depends on the nature of the data available.

It is important to highlight that the main critique of educational mismatch literature comes from its measurement methods. It is found that the incidence of educational mismatch generally differs depending on the measurement method considered, although a specific database is used that refers to the same period of time and country.

However, the effects of educational mismatch on different labour market outcomes, especially on wages, are quite robust regardless of the measurement method used (Hartog, 2000).

1.3. The impact of overeducation

Overeducation could have negative consequences on workers, firms and the economy as a whole.

Most of the literature on overeducation focuses on its effects on individual wages. Following the wage model specified by Duncan and Hoffman (1981), it is found that overeducated workers earn less than properly educated workers with the same educational level, but more than their properly educated work colleagues (Hartog, 2000; Groot and Maassen van der Brink, 2000; Rubb, 2003a). On the other hand, undereducated workers earn more than properly educated workers with the same level of education, but less than their properly educated work colleagues[1]. Sattinger and Hartog (2013) provide a theoretical foundation for those empirical regularities based on Nash bargaining between workers and employers.

Besides the wage penalty, overeducated workers have less job satisfaction (Verhaest and Omey, 2009; Korpi and Tåhlin, 2009), less life satisfaction (Piper, 2013) and more depression symptoms (Bracke, et al., 2013). They also have higher turnover rates (Alba-Ramirez, 1993; Sloane, et al., 1999; Quintini,

[1] Undereducation has not attracted the same attention of researchers, because undereducated workers do not have any incentives to leave a job that requires a higher level of education. Furthermore, firms could also have the benefit of hiring workers with an initial lack of education or knowledge if the cost of providing them specific job training is less than the benefit of giving them fewer wages than properly educated workers.

2011), an adverse impact on productivity (Tsang et al. 1991) but they experience greater promotion within firms (Groeneveld and Hartog, 2004).

On the other hand, firms run the risk of lower profits primarily because of lower overeducated worker effort (higher absenteeism, etc.) induced by low job satisfaction (Levin and Tsang, 1985). Belfield (2010) shows how overeducation rates across workplaces adversely influence workplace labor relations.

From a macroeconomic point of view, overeducation could be considered a waste of resources in economic terms given that high education is highly subsidised by governments and that they obtain fewer economic returns of investment in education from populations that end up overeducated.

1.4. Theoretical background

Some theories related to the labour market have been considered to explain the incidence of the overeducation phenomenon and its impact on wages.

The *human capital* theory considers that an individual's particular level of human capital will provide a certain level of productivity, irrespective of the job in which that individual works (Becker, 1975). It suggests that individuals are paid the value of their marginal product. Thus if one adopts a strict interpretation of human capital theory, where markets are efficient, one would not observe overeducation in the labour market other than as a short-term dynamic problem (Dolton and Vignoles, 1998). However, this characterisation assumes that firms are able to fully utilise the skills and knowledge of their employees and can adapt their production technology in response to changes in the relative supply of skilled workers. However, empirical analysis does not support this theory given that wages of workers with the same level of education vary depending on their occupation match (Hartog and Oosterbeek, 1988; Hartog, 2000; Leuven and Oosterbeek, 2011).

Another explanation consistent with the human capital theory is the *career mobility* theory (Sicherman and Galor, 1990). It considers overeducation as an investment opportunity. Sicherman and Galor (1990) defined a model that considers that a worker with a given level of education may prefer to start in a job below his or her educational level if this is compensated with a higher probability to be promoted. Under this point of view, overeducation is a transitory phenomenon explained by the worker's lack of experience. Although the finding that overeducation is more prominent among younger workers is consistent with this theory, there is empirical evidence that the probability of a job mismatch does not significantly decrease with experience (Frei and Sousa-Poza, 2012; Rubb, 2003b). In this sense, Baert et. al (2013) also find that once a youth worker accept a job for which is overeducated, monthly transitions rates into adequate employment fall by 51-98%.

On the other hand, *job competition* theory (Thurow, 1975) predicts that job characteristics are the only factor determining wages and qualifications are only important for the allocation of jobs. Individuals compete for job opportunities based on their relative training costs – i.e., their position in the hiring queue is determined by their cost in terms of training. This model assumes that qualifications serve as a proxy for training costs with the more highly qualified seen as more able and therefore requiring less training by the firm. Under this point of view, overeducation is a persistent phenomenon and that wages are fully dependant on required qualifications while returns to surplus qualifications are zero. Empirical studies do not support this theory given that the returns to surplus qualification are positive and statistically significant (Hartog and Oosterbeek, 1988; Hartog, 2000; Leuven and Oosterbeek, 2011).

The *assignment* theory (Sattinger, 1993) is a model between human capital theory on the one hand and job competition on the other. It considers that workers' productivity is limited by their workplace and in part by their human capital. The basic idea is that, although education raises productivity in general, working in a job below one's own qualification level imposes a ceiling on a worker's productivity because it limits the extent to which his or her skills can be utilised and results in lower wages. According to assignment theory, productivity is maximised when workers are allocated top-down according to their skills, whereby the most skilled are assigned to the most complex job and the least skilled to the simplest job. Overeducation is explained by differences in the share of complex jobs and skilled workers. Although it has been one of the most supported theories, different empirical analyses that have tried to test it do not support it (Allen and van der Velden, 2001; Di Pietro and Urwin, 2006; Green and McIntosh, 2007; Sánchez-Sánchez and McGuinness, 2013; Mavromaras et al. 2013).

Finally, the *heterogeneous skills* theory (Green and McIntosh, 2007) considers that there is significant variability in terms of skills among individuals with the same level of education. Thus it is quite possible to find workers who appear to be overeducated, but their skills level may match more closely that of those with the appropriate level of education for the job they occupy. Accordingly, the reason why overeducated workers are found to earn less than their peers with the same level of education who work in jobs for which their qualifications are appropriate is because the former are either less able or have fewer skills. This theory seems to be the most supported by recent literature given that it is found that the effect of overeducation on wages reduces or even disappears when it is controlled for all unobserved individual fixed effects (Bauer, 2002; Frenette, 2004; Korpi and Tåhlin, 2009; Tsai, 2010).

1.5. Three empirical studies on overeducation

This section presents a summary of the three empirical analyses on overeducation in Spain and their contribution to the literature.

Chapter 2: "Non-formal education, overeducation and wages"

Although there is an important incidence of overeducation, demand for skills is growing in most countries. In this sense, important efforts are being devoted to promote lifelong learning, since it permits both individuals and society to better adapt to changes in economic conditions.

In spite of the fact that overeducated workers have more attained education than that required by their jobs, it is found that overeducated workers participate more in training activities in comparison with their adequately educated colleagues working at the same job level, but they get significantly less training in comparison with adequately educated individuals with the same level of education (Verhaest and Omey, 2006).

The objective of this chapter is twofold. First, it will examine the effect of non-formal education activities on individual wages and test whether this effect is different depending on workers' years of schooling. And, second, the chapter will analyse if the returns of the participation in these types of training activities is higher for overeducated workers than for the rest of workers. If it is so, overeducated workers could overcome part of the wage penalty derived from their education-occupation mismatch. The results show that non-formal education activities have a positive effect on wages, this effect being higher for those workers with higher education. Moreover, overeducated workers suffer a wage penalisation compared to well-educated workers with the same level of education. However, only overeducated workers who have undergone non-formal education activities receive a wage premium. It seems that this type of training seems to provide overeducated workers with new abilities that permit them to reduce the wage penalisation derived from the mismatch between their level of education and occupation.

The contribution of this chapter is to show a way to modify the negative consequences of overeducation on wages. Overeducated workers are the ones that benefit from getting training, because it allows them to reduce the wage penalty associated with their overeducation.

Chapter 3: Overeducation, skills and wage penalty

A supported theory on overeducation is based on the existence of individuals' skill heterogeneity. From such a perspective, the wage penalty associated with overeducation is due to the huge variation of skills among workers with the same level of education. Then, overeducated workers would not suffer a wage penalty. In fact, they would earn lower wages as a result of their lower level of skills. However, most of the literature does not explicitly test this hypothesis due to data limitations regarding individuals' skill levels.

The specific aim of the chapter is to test the skill heterogeneity theory in Spain. Our hypothesis is that the wage penalty associated with overeducation could be explained by lower skill levels. In consequence, overeducated workers may not be suffering a wage penalty in Spain, but their earnings are determined by their skill level.

Our results show that individuals' skill heterogeneity only explains 18% of the effect of educational mismatch on wages in Spain. The wage penalty still remains for those overeducated workers who are not less skilled than properly matched workers.

This chapter is a contribution to the literature of overeducation that takes into account workers' individual skills using direct indicators of skill levels. In the case of Spain, workers' skill heterogeneity explains little about the wage loss of overeducated workers.

<u>Chapter 4: Is there a link between parents' overeducation and the students' educational achievement?</u>

As Haveman and Wolfe (1995) highlight, children of highly educated parents tend to perform better than children of less educated parents. One possible explanation for the positive relationship between parents' human capital and students' performance is based on children's perceptions about the importance of education. In this sense, students whose parents have a high level of education and good jobs might be more aware of the value of education and, consequently, have higher motivation and perform better than other students. Under this point of view, our hypothesis is that the existence of parents' job-education mismatch can modify the students' perception about the importance of education and, consequently, have an effect on their performance at school.

In particular, the objectives in this chapter are the following: first, we analyse whether there is a relationship between parents' educational mismatch and the educational performance of their children. And, second, in case this relationship exists, we check whether it is similar across the performance distribution or, by contrast, whether there are differences between students at the top and at the bottom of the performance distribution.

The results show a statistically significant relationship between parents' educational mismatch and children's educational performance after controlling for the effect of individual and school characteristics and other family-related variables. On the one hand, students whose parents are overeducated have a penalty in their academic achievement in all three subjects analysed, this effect being stronger for students with lower educational outcomes. On the other hand, undereducation only affects students' educational achievement when it is suffered by the mother, this effect being positive. So, the results confirm our hypothesis, although they cannot prove that the students' perception is the transmission channel, an aspect that is beyond the scope of the current research due to data limitations.

This chapter makes a contribution by identifying a new collective that can be affected by the overeducation phenomenon. Whilst previous literature focuses on the effect of overeducation on workers, we identify that there could also be an impact of parents' overeducation on the academic performance of their children.

References

Alba-Ramirez, A. (1993) "Mismatch in the Spanish labor market: overeducation?" *Journal of Human Resources*, 28 (2), 259-278.

Allen, J. and Van der Velden, R. (2001) "Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search", *Oxford Economic Papers*, 53(3), 434-452.

Baert, S.; Cockx, B. and Verhaest, D. (2013) "Overeducation at the start of the career: stepping stone or trap?", *Labour Economics,* 25, 123-140.

Bauer, T. (2002) "Educational Mismatch and Wages: A Panel Analysis", *Economics of Education Review*, 21, 221–229.

Becker, G. S. (1975) *Human Capital*, 2nd edn, National Bureau of Economic Research (NBER), New York.

Belfield, C. (2010) "Over-education: What influence does the workplace have?", *Economics of Education Review*, 29(2), 236-245.

Bracke, P.; Pattyn, E. and von dem Knesebeck, O. (2013) "Overeducation and depressive symptoms: Diminishing mental health returns to education", *Sociology of Heath and Illness,* 35(8), 1242-1259.

Di Pietro, G. and Urwin, P. (2006) "Education and skills mismatch in the Italian graduate labour market", *Applied Economics*, 38(1), 79-93.

Dolton, P. and Vignoles, A. (1998) "Overeducation: problem or not?" in *Changing Relationship between Higher Education and the State* (Eds) M. Henkel and B. Little, Jessica Kingsley, London.

Duncan, G. and Hoffman, S. (1981) "The incidence and wage effects of overeducation", *Economics of Education Review*, 1(1), 75-86.

Freeman, R. (1976) *The Overeducated American*. Academic Press.

Frei, C. and Sousa-Poza, A. (2012) "Overqualification: permanent or transitory?", *Applied Economics*, 44(14), 1837-1847.

Frenette, M. (2004) "The overqualified Canadian graduate: the role of the academic program in the incidence, persistence, and economic returns to overqualification", *Economics of Education Review*, 23, 29–45.

Green, F. and McIntosh, S. (2007) "Is There a Genuine Under-utilisation of Skills Amongst the Over-qualified?", *Applied Economics*, 39(4), 427-439.

Groeneveld, S. and Hartog, J. (2004) "Overeducation, Wages and Promotions within the Firm", *Labour Economics*, 11, 701-714.

Groot, W. and Maassen van den Brink, H. (2000) "Overeducation in the Labor Market: a Meta-Analysis", *Economics of Education Review*, 19(2), 149-158.

Hartog, J. (2000) "Over-education and earnings: Where are we, where should we go?", *Economics of Education Review*, 19, 131-147.

Hartog, J. and Oosterbeek, H. (1988) "Education, allocation and earnings in the Netherlands: overschooling?", *Economics of Education Review*, 7(2), 185–194.

Haveman, R. and Wolfe, B. (1995) "The determinants of children's attainment: A review of methods and findings", *Journal of Economics Literature*, 33(4), 1829-1878.

Kiker, B.F.; Santos, M.C. and Mendes de Oliveira, M. (1997) "Overeducation and Undereducation: Evidence for Portugal", *Economics of Education Review*, 16(2), 111-125.

Korpi, T. and Tåhlin, M. (2009) "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000", *Labour Economics*, 16, 183-193.

Leuven, E. and Oosterbeek, H. (2011) "Overeducation and Mismatch in the Labor Market", in: E. Hanushek, S. Machin, L. Woessman (Eds.), *Handbook of the Economics of Education*, 4, 283-326.

Levin, H. M., and Tsang, M. (1985) "The economics of overeducation", *Economics of Education Review*, 4, 93-104.

Mas, M.; Perez, F.; Uriel, E. and Serrano, L. (1995) *Capital Humano Series Históricas, 1964-1992*, Fundación Bancaixa.

Mavromaras, K.; McGuinness, S.; O'Leary, N.; Sloane, P. and Wei, Z. (2013) "Job Mismatches and Labour Market Outcomes: Panel Evidence on University Graduates", *Economic Record*, 89, 382–395.

McGuinness, S. (2006) "Overeducation in the Labour Market," *Journal of Economic Surveys*, 20(3), 387-418.

OECD (2009) *Education at a glance 2009*, Paris.

OECD (2013) *OECD Skills Outlook 2013 First Results from the Survey of Adult Skills*, Paris.

Piper, A. (2014) "Heaven knows I'm miserable now: overeducation and reduced life satisfaction", *Education Economics*, forthcoming.

Quintini, G. (2011) "Over-Qualified or Under-Skilled: A Review of Existing Literature", *OECD Social, Employment and Migration Working Papers,* No. 121, OECD Publishing.

Rubb, S. (2003a) "Overeducation in the labor market: A comment and re-analysis of a meta-analysis", *Economics of Education Review*, 22, 621-629.

Rubb, S. (2003b) "Overeducation: A short or long run phenomenon for individuals?", *Economics of Education Review*, 22, 389-394.

Sánchez-Sánchez, N. and McGuinness, S. (2013) "Decomposing the impacts of overeducation and overskilling on earnings and job satisfaction: an analysis using REFLEX data", *Education Economics*, forthcoming.

Sattinger, M. (1993) "Assignment Models of the Distribution of Earnings", *Journal of Economic Literature*, 31, 831-880.

Sattinger, M. and Hartog, J. (2013) "Nash bargaining and the wage consequences of educational mismatches", *Labour Economics*, 23(1), 50-56.

Sicherman, N. and Galor, O. (1990) "A theory of career mobility", *Journal of Political economy*, 98(1), 169–192.

Sloane, P.; Battu, H. and Seaman, P. (1999) "Overeducation, Undereducation and the British Labour Market," *Applied Economics*, 31, 1437-1453.

Thurow, L. (1975) *Generating inequality*. Basic Books.

Tsai, Y. (2010) "Returns to overeducation: A longitudinal analysis of the U.S. labor market", *Economics of Education Review*, 29, 606-617.

Tsang, M. C.; Rumberger, R. and Levin, H.M. (1991) "The impact of surplus schooling on worker productivity", *Industrial Relations*, 30(2), 209-228.

Verdugo, R. and Verdugo, N. (1989) "The impact of surplus schooling on earnings", *Journal of Human Resources*, 24(4), 629-643.

Verhaest, D. and Omey, E. (2006) "The impact of overeducation and its measurement", *Social Indicators Research*, 77, 419-448.

Verhaest, D. and Omey, E. (2009) "Objective over-education and worker well-being: A shadow price approach," *Journal of Economic Psychology*, 30(3), 469-481.

Verhaest, D. and van der Velden, R. (2013) "Cross-country Differences in Graduate Overeducation", *European Sociological Review*, 29(3), 642-653.

Chapter 2: Non-formal education, overeducation and wages[2]

2.1. Introduction

Traditionally, there has been a clear separation between two different ways of accumulating human capital: schooling at an early age and on-the-job experience in adulthood. Nowadays, this separation is not so clear. The role of lifelong learning as a way through which individuals can accumulate human capital beyond early adulthood is a central issue in the current European education policy. In fact, while recognising the role of primary, secondary and higher education, the "Strategic framework for European cooperation in education and training – ET2020" gives priority to lifelong learning as a way to adapt to a rapidly changing world.

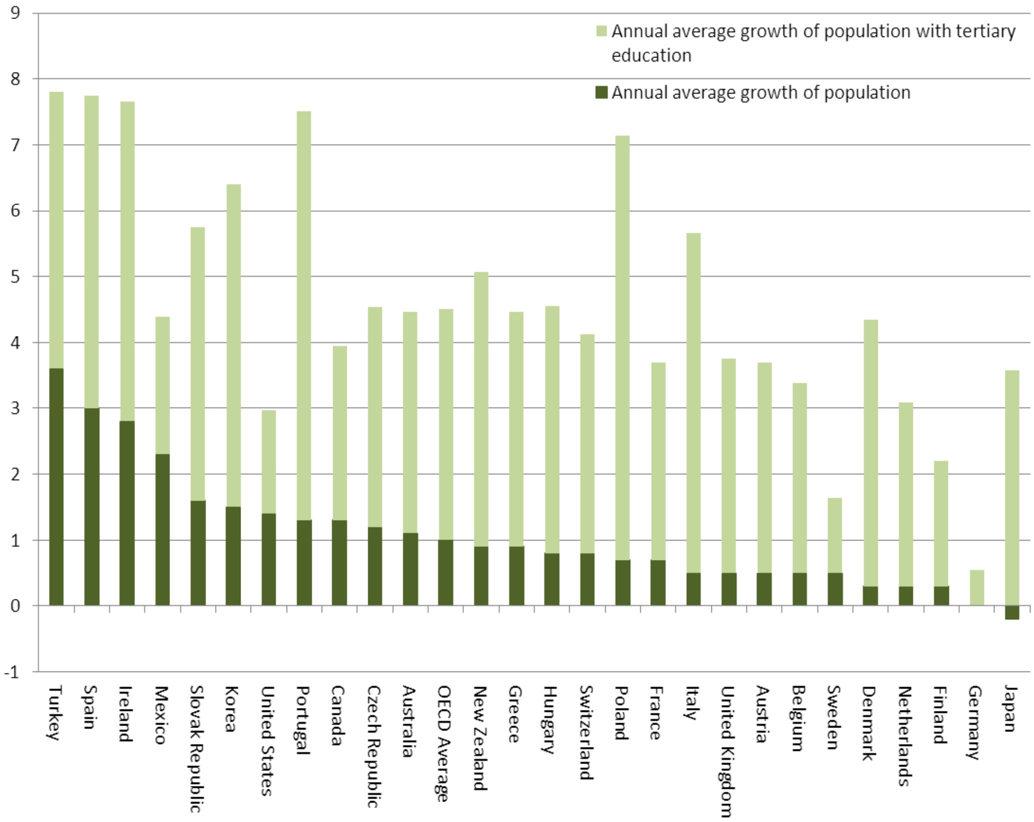
From the seminal contribution of Mincer (1974), there is a consensus about the positive effect of education on wages. That is, high-educated workers earn higher wages than workers with lower educational levels. However, returns on education have decreased over time. This was indeed pointed out for the U.S. by Freeman (1976) in the book *The Overeducated American*. He explains that this trend may be the consequence of an excess supply of high-educated workers

[2] Previous versions of this research have been presented in XIV Encuentro de Economía Aplicada 2011 (Huelva, Spain); XX Jornadas de la Asociación de Economía de la Educación 2011 (Málaga, Spain); IX Jornadas de Economía Laboral (Santiago de Compostela, Spain); IV Congreso Nacional sobre Mercado de Trabajo y Relaciones Laborales (Palencia, Spain). It has also been published as "Nieto, S. and Ramos, R. (2013) Non-formal education, overeducation and wages, Revista de Economía Aplicada, 61(XXI), 5-28."

caused by an overinvestment in college education in the U.S. On the other hand, analyses of the effect of lifelong learning activities on wages in later adulthood are scarce and less conclusive. Furthermore, they mainly focus on formal education in adulthood as lifelong learning activity, while other non-formal education activities have not been considered. For instance, Egerton and Parry (2001) report substantial penalties for late learners since the returns on these activities are fewer than their estimated costs. Jenkins et al. (2002), on the other hand, find little evidence of the positive effect of lifelong learning on wages. In particular, they find positive and significant earning returns for men who left school with low-level qualifications and subsequently obtained a degree. De Coulon and Vignoles (2008) reveal a positive effect of 18% on earnings (on average) of lifelong learning, while Blanden et al. (2012) provide evidence of a medium-run return for women of 10% on hourly wages.

A trend in education similar to the one observed by Freeman (1976) seems to appear in most of the OECD countries, since they have experienced an important increase in the number of graduates (see Figure 2.1). In particular, the annual average growth in graduates across all the OECD countries between 1998 and 2006 was around 4.5%, far above the annual population growth in the same period, which was around 1%. The case of Spain deserves special attention, because, in the 1960s and 1970s, Spain's population had a very low level of education in comparison with the other OECD countries (Mas et al., 1995). Over the last decades, Spain has experienced an annual average growth in the number of graduates of almost 8%, accompanied by an average population increase of approximately 3.5%.

Figure 2.1. Annual average growth in 25-64 year-old population between 1998 and 2006



Source: OECD (2009)

This sharp increase in the number of graduates has led to a situation in which the percentage of highly educated workers in Spain is currently above the OECD average, although the occupational structure of the Spanish economy is clearly dominated by low and medium skill jobs. As a result, Spain is one of the developed countries with the highest incidence of overeducation, i.e., workers who have a level of education higher than that required by their jobs. According to the OECD (2007), overeducated workers accounted for 25% of total employees in Spain in 2004, more than double the OECD average (12%). Then, one in every four workers in Spain suffers the negative effects of overeducation, especially on their wages. In particular, it is found that overeducated workers receive smaller wages than properly educated workers with the same educational level (see for a review Hartog, 2000; and Leuven and Oosterbeek, 2011).

García-Montalvo (2005) points out that the higher incidence of overeducation could be related to an excessive supply of highly educated individuals that the job market has been incapable of absorbing because the jobs on offer do not require such a high level of education, but also to the lack of practical competences of the graduates. In this sense, Sloane (2003) also argues that workers could be overeducated because they do not have the required skills and competences to perform the job satisfactorily and that these skills probably could not be acquired through formal education, an argument that reinforces the role of lifelong learning through activities different from formal education.

Regarding the relation between overeducation and lifelong learning activities, Verhaest and Omey (2006) and Büchel (2002) find that overeducated workers participate more in training activities in comparison with their adequately educated colleagues working at the same job level, but they acquire significantly less training in comparison with adequately educated individuals with the same level of education (Hersch, 1991; van Smoorenburg and van der Velden, 2000; Büchel and Mertens, 2004; and Verhaest and Omey, 2006). This finding suggests that, first, more educated workers get more training activities, an idea that is related to the fact that education and training activities are complements (Rosen, 1976). And, second, overeducated workers may not perform non-formal education activities because of a lack of skills, since they tend to carry out these types of activities less frequently than workers with the same level of education who are well matched in their jobs.

Taking into account these previous considerations, the objective of this chapter is twofold. First, its objective is to examine the effect of non-formal education activities on individual wages and test whether this effect is different depending on workers' years of schooling. And, second, its objective is to determine if the returns of the participation in these types of training activities is higher for overeducated workers than for the rest of workers. If it is so, overeducated workers could overcome part of the wage penalisation derived from their education-occupation mismatch. The analysis is carried out using microdata from the Spanish sample of the 2007 Adult Education Survey, a survey that provides detailed information on lifelong learning among the adult population.

According to our results, the participation in non-formal education activities has a positive effect on individual wages. We have also found that non-formal education seems to provide overeducated workers with new abilities that permit them to reduce the wage penalisation derived from their skill mismatch.

The rest of chapter is structured as follows: Section 2.2 presents the data and the variables used in the analysis. It also explains how educational mismatch is measured. Next, Section 2.3 shows a descriptive analysis of the relation between non-formal education activities and educational mismatch. Section 2.4 describes the methodology strategy. Section 2.5 presents the results of estimating the return of the different types of human capital considered in the study, paying special attention to the interaction between non-formal education and educational mismatch. Finally, Section 2.6 contains some closing remarks.

2.2. Database and variables

2.2.1. The Spanish sample of the 2007 Adult Education Survey

The most appropriate survey for analysing lifelong learning among adults is the Adult Education Survey (AES). The main objective of the survey is to study lifelong learning – that is, those training and learning activities that the adult population performs with the objective of improving or extending their knowledge, skills and competences from a personal, civil, social or work-related perspective. In this work, we use the Spanish sample of the 2007 AES[3] *(Encuesta sobre la Participación de la Población Adulta en las Actividades de Aprendizaje; EADA*) that was carried out by the Spanish National Institute of Statistics (INE). This survey provides information about a sample of 20,009 people aged between 25 and 74 from all over Spain.

[3] Now AES 2011 is available. However, this wave was not available when this research was conducted.

The learning activities considered by the AES are the following:

- Formal education: it includes education provided in the system of schools, colleges, universities and other formal educational institutions that normally constitutes a continuous "ladder" of full-time education for children and young people, generally beginning at the age of 5 to 7 and continuing up to 20 or 25 years old. It leads to recognised qualifications.

- Non-formal education (NFE): it refers to institutionalised learning activities, which are not considered formal education. It includes structured programmes that cannot be positioned in the ISCED usually because of the provider and/or the awarding organisation, content or duration. For example, language courses, computer training or art classes are considered non-formal education activities.

- Informal education: it considers activities that are carried out with the intention of learning, but are not as organised or structured as educational activities. It includes activities such as reading the newspaper or looking for specific information in a book or on the internet.

We restrict our analysis to non-formal education activities since they are intended to improve the population's specific skills, but without modifying the amount of attained years of schooling.

It is worth mentioning that this is the only survey carried out in Spain that provides a high degree of detail on formal and non-formal education. Previous studies focusing on the analysis of educational mismatch and its impact on wages have used microdata extracted from different surveys, such as: the *Encuesta de Calidad de Vida en el Trabajo* (Quality of Life at Work Survey; ECTV), the *Encuesta de Población Activa* (Labour Force Survey; LFS), the *Encuesta de Estructura, Conciencia y Biografía de Clase* (Class Biography, Conscience and Structure Survey; ECBC), the European Community Household Panel (ECHP), the *Encuesta de Presupuestos Familiares* (Household Budget Survey;

EPF) or the *Encuesta de Estructura Salarial* (Structure of Earnings Survey; SES). However, these sources do not provide information on lifelong learning, with the exception of the LFS, which devotes seven questions to training activities but provides very little information on non-formal education activities. For this reason, we choose to use microdata from the AES survey.

Beyond information about lifelong learning activities, the survey provides variables related to personal and job characteristics. With respect to the personal characteristic variables, we use information related to gender, nationality, years of education[4] and the number of household members. As for job characteristics, we consider workers' occupation, monthly earnings, economic activity, potential experience (age minus the number of years in education minus six), seniority, type of contract, type of working day, number of jobs and firm size. Finally, we also consider information about the region (Autonomous Communities) and regional population density.

Moreover, although AES is a dataset containing cross-sectional data referring to 2007, it provides retrospective information on a series of variables related to the occupation of workers corresponding to the 12 months preceding the survey. Consequently, it also allows us to measure educational mismatch in 2006 beyond 2007.

The initial sample considered is composed of people who were working at the time of the survey (11.738 observations). However, the sample is restricted to 5.183 observations when the effect of non-formal education and educational mismatch on wages is estimated due to an important number of workers having missing values in the variables related to jobs (mainly wages). Descriptive statistics for both samples are shown in Table A2.2 of the Annex. As we can see, there are no important differences between both samples.

[4] The AES provides data on schooling levels. The equivalences applied to calculate the number of schooling years are shown in Table A2.1 of the Annex.

2.2.2. Measurement of educational mismatch

As it has been explained in the section 1.2 of chapter 1, literature on educational mismatch has proposed (at least) three different methods for measuring educational mismatch.

The AES database allows us to measure educational mismatch only through the statistical method (both the version based on the mean and the version based on the mode) given that it informs workers' attained education and their occupations[5]. In particular, the statistical method based on the mean (Verdugo and Verdugo, 1989) considers that a person is overeducated (undereducated) if he/she has a level of education that is higher (lower) by more than one standard deviation than the mean level of education of the workers in that occupation. Nevertheless, Kiker et al. (1997) propose the use of the mode instead of the mean; so they consider a person who has a higher (lower) level of education than the mode for the job they perform to be overeducated (undereducated).

Following both definitions, Table 2.1 shows the percentages of educational mismatch for 2006 and 2007. As we can see, the incidence of educational mismatch varies according to the procedure used. In particular, the method based on the mean shows that undereducated workers represent 21% of the total number of workers in 2007, while almost 20% are overeducated. However, using the method based on the mode, both overeducation and undereducation are greater. Specifically, the percentage of undereducated workers is almost 32% of all workers, while overeducated workers are 24%, more than 10 and 5 percentage points (respectively) above the figure obtained using the method based on the mean. As is to be expected, the results for 2006 are very similar to those for 2007.

[5] Two digits disaggregation.

Table 2.1. Educational mismatch in 2006 and 2007 (in percentages)

	Mode		Mean	
	2006	2007	2006	2007
Undereducation	31.62	31.98	21.52	21.05
Required education	43.37	43.81	58.54	59.41
Overeducation	25.01	24.21	19.94	19.54
Total	100	100	100	100

This finding is probably a consequence of the definition of educational mismatch adopted in each of the methods. Specifically, the method based on the mean classifies workers as properly educated if they are within one standard deviation of the mean number of years of education that people in a specific occupation have completed, whereas the version based on the mode counts workers as properly educated only when the number of years of education they have completed coincides exactly with the mode. This means that a worker is more likely to be considered properly educated in the version based on the mean than in the version based on the mode. This result is, indeed, a common finding in educational mismatch literature (see Table 2.2).

In comparison to the subjective and the objective methods, Bauer (2002) observes that the level of educational mismatch given by the statistical method is less than that given by either the objective or subjective method and, therefore, that the statistical method is considered to underestimate educational mismatch. In fact, the statistical method takes the ideal level of education for performing a specific occupation to correspond to the mean (mode) number of years of formal education that people who work in that occupation received. Consequently, if the majority of workers in a specific occupation were overeducated, the mean (mode) for that occupation would be higher than it would be if the majority of the workers were not overeducated; therefore, in the former case, overeducation would be underestimated. Analogously, if the majority of workers were undereducated, the mean (mode) would be lowered and undereducation would be underestimated.

Table 2.2. Survey of research on educational mismatch in Spain

Author	Data source and year	Method	Overed.	Undered.
Alba-Ramírez (1993)	ECVT, 1985	Subjective (indirect)	17.0	23.0
García-Montalvo (1995)	EPA, 1985	Objective	3.7	30.4
	EPA, 1987	Objective	5.1	31.2
	EPA, 1989	Objective	6.3	31.0
	EPA, 1991	Objective	6.6	30.5
	EPA, 1993	Objective	7.7	27.6
	EPA, 1993	Statistic (mean)	8.9	6.2
García Serrano and Malo (1996)	ECBC, 1991	Subjective (indirect)	28.4	30.0
	ECBC, 1991	Subjective (direct)	29.4	11.0
Beneito et al. (1996)	ECBC, 1991	Statistic (mean)	15.2	15.3
Oliver and Raymond (2002)	EPA, 2001	Statistic (mode)	32.5	n.a.
Badillo-Amador et al. (2005)	ECHP, 1998	Subjective (direct)	34	44
		Statistic (mode)	35	26
Aguilar and García (2008)	ECHP, 1995	Subjective (direct)	53.4	24.6
	ECHP, 1998	Subjective (direct)	58.8	19.7
	ECHP, 2001	Subjective (direct)	60.9	20.9
Barone and Ortiz (2011)	REFLEX, 2005 (only graduates)	Subjective (indirect)	17.1	n.a.
Quintini (2011)	ESWC	Statistic (mode)	32	31
Murillo et al. (2012)	EES, 1995	Statistic (mode)	35.3	20.8
	EES, 2002	Statistic (mode)	31.9	25.6
	EES, 2006	Statistic (mode)	37.2	23.0
Ramos and Sanromá (2013)	EPF, 1990-91	Statistic (mean)	14.6	n.a.

Note: n.a. = not available

However, it is important to highlight that in spite of the differences in percentages depending on the measurement method considered, Hartog (2000) concluded that the basic relation between educational mismatch and earnings seems not to be influenced by the measurement method. Nevertheless, Rubb (2003) points out that the measurement influences the magnitude of the coefficients.

According to this, we perform the main analysis using one of the measurement methods – in particular, the mode version[6]. However, we do a robustness check using the mean version (see Section 2.5.3).

2.3. Descriptive analysis

First, using the information on educational mismatch for 2006 and 2007, we analyse whether workers who have participated in non-formal education activities change their situation of education-occupation mismatch in a different way than the rest of workers (Table 2.3). If this is so, non-formal education activities could encourage workers to change their occupation or, alternatively, help them to gain promotion. In this sense, it is important to highlight that non-formal education consists of those activities that do not lead to a recognised qualification and that people of all ages can embark on (so, by itself it does not change the situation of education mismatch of workers). This type of activity can lead to an increase in the competences and skills of those who undertake them. However, non-formal education is not counted when determining the number of years of education a person has received. Thus, in this study, people's level of education is considered to be given by the number of years of formal education they have received, which remains unaltered by participation in non-formal education, although the latter is considered a means of increasing competences and skills.

[6] Hartog (2000) explains that the method based on the mean tends to produce very similar percentages of overeducation and undereducation as a consequence of the symmetry of the tails of the normal distribution.

Table 2.3. Evolution of the educational mismatch from 2006 to 2007

	Workers		Workers who have undertaken NFE activities	
	Frequency	%	Frequency	%
Mismatch not changed	10.994	94.86%	3.625	93.67%
Mismatch changed	596	5.14%	245	6.33%
Total	11.590	100%	3.870	100%
Mismatch changed				
From properly to over	57	9.56%	24	9.80%
From properly to under	203	34.06%	76	31.02%
From over to properly	253	42.45%	115	46.94%
Other situations	83	13.93%	30	12.24%
Total	596	100%	245	100%

Note: NFE=non-formal education. Common sample of employees is reduced to 11.590 in both years.

Table 2.3 shows that approximately 5% of all workers and 6% of workers who undertook non-formal education changed the match between their level of education and their occupation between 2006 and 2007. Moreover, more than 42% of the people whose degree of educational mismatch changed managed to equate their level of education to that required for the work they perform. However, the change of occupation or internal promotion does not seem to have been due to training activities, since this change can be observed in both those who undertook this type of activity and those who did not. Thus, the fact that both samples present a low percentage of workers who have changed their type of match between the level of education and occupation seems to indicate that undertaking non-formal education activities does not have a great impact on the type of mismatch.

On the other hand, we also analyse the participation of workers in non-formal educational activities depending on their education-occupation match and educational level (Table 2.4). First, it can be seen that 46.1% of workers with tertiary education have undertaken non-formal education activities, while the percentage for workers with primary education is 19.2%. The results support the complementarity between education and training. On the other hand, Table 2.4 also shows that only the 23% of undereducated workers, who are those with less formal education than their job requires, realise non-formal education activities. In contrast, the 38% of overeducated workers have participated in some kind of non-formal education, while this percentage is 35% for workers without an education-occupation mismatch.

Furthermore, if we focus on workers with tertiary and secondary education, overeducated workers participate more in training activities in comparison with their adequately educated colleagues working at the same job level, but they get less training in comparison with adequately educated individuals with the same level of education. Our results are in line with previous literature (Verhaest and Omey, 2006; Büchel, 2002; Hersch, 1991; van Smoorenburg and van der Velden, 2000; Büchel and Mertens, 2004; and Verhaest and Omey, 2006).

Table 2.4. Percentage of workers who have undertaken NFE activities in the last 12 months

	Primary	Secondary	Tertiary	Total
Undereducated	19.7	35.7	58.4	23.2
Properly educated	17.7	35.8	46.4	35.1
Overeducated	20.1	28.7	44.1	37.9
Total	19.2	34.2	46.1	

Note: NFE=non-formal education

In this sense, if overeducated workers are in jobs that require less education than they already have and non-formal education does not seem to help them to improve their situation and obtain a proper job, why is their participation in non-formal education activities higher than it is for their adequately educated colleagues working at the same job level? In the following sections, we analyse whether overeducated workers may receive a wage premium compared to other overeducated workers in order to explain their higher participation in these kinds of learning activities or, on the other hand, if this only reflects the fact that high-educated workers get more training than workers with a lower educational level.

2.4. Methodology and results

Firstly, in order to analyse the effect of non-formal education activities on individual wages and test whether this effect is different depending on workers' schooling years, we estimate the following specifications of the Mincer wage equation (Mincer, 1974):

$$\log(W_i) = \alpha + \beta' \cdot X_i + \delta \cdot S_i + \rho_1 \cdot E_i + \rho_2 \cdot E_i^2 + u_i \quad (1)$$

$$\log(W_i) = \alpha + \beta' \cdot X_i + \delta \cdot S_i + \rho_1 \cdot E_i + \rho_2 \cdot E_i^2 + \rho_3 \cdot NFE_i + u_i \quad (2)$$

$$\log(W_i) = \alpha + \beta' \cdot X_i + \delta \cdot S_i + \rho_1 \cdot E_i + \rho_2 \cdot E_i^2 + \rho_3 \cdot NFE_i + \rho_4 \cdot S\_NFE_i + u_i \quad (3)$$

where $log(W_i)$ represents the logarithm of the monthly wage of worker i; X_i is a vector of variables related to personal characteristics and employment characteristics; S_i refers to the number of years of formal education; E_i represents experience; E_i^2 represents the square of experience; NFE_i is a dummy variable that takes the value 1 if the worker has partaken in any non-formal education activity and 0 otherwise; S_NFE_i is defined as the interaction of variable S with NFE; and, finally, u_i is the error term with zero mean and constant variance[7].

Secondly, we estimate an ORU (Over-Required-Undereducated) by Duncan and Hoffman (1981) in order to analyse the effect of overeducation on wages. In fact, the ORU specification is a variant of the Mincer equation in which the years of education of the worker (S) are separated into years of education required for the job (Sr), years of overeducation (So) and years of undereducation (Su). More specifically, So and Su are defined as:

$$So = \begin{cases} S - Sr & if \quad S > Sr \\ 0 & on \quad contrary \end{cases} \qquad Su = \begin{cases} Sr - S & if \quad S < Sr \\ 0 & on \quad contrary \end{cases}$$

The ORU equation is then defined as:

[7] Mincer equations have been estimated using ordinary least squares that are robust to heteroscedasticity.

$$\log(W_i) = \alpha + \beta' \cdot X_i + \delta_1 \cdot Sr_i + \delta_2 \cdot So_i + \delta_3 \cdot Su_i + \rho_1 \cdot E_i + \rho_2 \cdot E_i^2 + u_i \tag{4}$$

Just as in the Mincer equation, X_i is a vector of variables related to personal and job characteristics, E_i represents potential experience and u_i is the error term.

Furthermore, in order to analyse whether non-formal education activities could report higher returns of workers depending on whether they are overeducated, undereducated or properly educated, we also estimate equation (5) including interactions between variables related to educational mismatch and non-formal education activities (Sr_NFE, So_NFE, Su_NFE_i):

$$\begin{aligned}\log(W_i) = {} & \alpha + \beta' \cdot X_i + \delta_1 \cdot Sr_i + \delta_2 \cdot So_i + \delta_3 \cdot Su_i + \rho_1 \cdot E_i + \rho_2 \cdot E_i^2 + \rho_3 \cdot NFE_i + \\ & \rho_4 \cdot Sr\_NFE_i + \rho_5 \cdot So\_NFE_i + \rho_6 \cdot Su\_NFE_i + u_i\end{aligned} \tag{5}$$

Finally, we also take into account that the results of each specification could mask a problem of sample selection bias or "self-selection" since the sample of workers with reported wages may no longer be random and may not be a suitable representation of workers overall – that is, when it comes to estimating the wage equation, only the characteristics of people who are working, which could lead to false conclusions about the effect of the different variables on the endogenous variable. For example, if the majority of workers in the sample considered have many years of education, there would be a tendency to underestimate the effect of education on wages.

To correct the possible selection bias, Heckman (1979) proposed a two-stage method that can be interpreted as the incorporation of variables that were omitted from the wage equation. The first stage consists of analysing the factors that determine the probability of being employed, from which Heckman's lambda is obtained (the inverse of Mills' ratio). In the second stage, Heckman's lambda is included as an explanatory variable in the wage equation. In this way, when the factors that determine wages are analysed, the uncontrolled characteristics relating to the probability of a person being employed are taken into account. Because of issues of identification, it is necessary that at least one explanatory variable that appears in the selection equation does not appear in the wage equation. In other words, we need a variable that affects the probability of being employed, but not the salary.

2.5. Results

In this section, we show the results of the Mincer and ORU equations to check the effect of non-formal education activities on individual wages and whether overeducated workers who perform these types of activities achieve higher returns than overeducated workers who do not. All estimations include the correction term of a possible problem of sample selection and individual weights.

2.5.1. Sample selection bias

As has been explained in the previous section, we apply the two-stage method proposed by Heckman (1979) in order to correct the possible selection bias.

In the first stage, we analyse the determinants of being employed considering that one of the variables, at least, should not be a determinant of workers' wages. Like in other studies, we consider the "number of household members" as an exclusion variable, as it can affect the decision to work but does not affect wages. In particular, people can decide to participate in the labour market depending on how many people they have in their charge, but we do not expect employers to consider this variable when setting their wages. The results of the first stage of the method are shown in Table 2.5. As we can see, women and immigrants are less likely to be employed than men and natives. Schooling years as non-formal education activities increase the probability of being in work. In contrast, the number of household members has a negative and statistically significant effect on the probability of being employed.

Table 2.5. Determinants of the probability of being in work

	(1)
Woman	-0.7899***
	[0.0224]
Immigrant	-0.1708***
	[0.0471]
Schooling years	0.0316***
	[0.0036]
Household members	-0.0393***
	[0.0088]
Experience	0.0611***
	[0.0034]
Experience squared	-0.0017***
	[0.0001]
Non-formal education	0.4055***
	[0.0267]
Urban size	Yes
Regional dummies	Yes
Constant	0.3005***
	[0.0835]
Observations	19989
Pseudo R2	0.3306

Note: Estimated coefficients. Standard errors in parentheses. *** statistically significant at 1%; ** statistically significant at 5%; * statistically significant at 10%. Individual sample weights are considered.

We have also tested the quality and the validity of this instrumental variable through common statistics. In particular, instrumental quality is ensured if there is a strong correlation between the instrument and the probability of employment; to test the joint significance, we have used the criteria suggested by Bound et al. (1995). The partial R squared and F statistic on the excluded instrument in the first-stage regression (using OLS) will indicate that the instrument is legitimate[8]. We have also checked the validity of the instrument through the approach of Dolton and Vignoles (2002): a valid instrument must be uncorrelated with the error term of the wage equation and, thus, it will not affect the income conditional on the included explanatory variables. Hence, we have regressed the residuals from the wage equation against the instrument,

[8] The partial R squared is 0.0117, and the value of the F statistic is 218.53.

and we have obtained an R squared of 0.0001, which means that the instrument does not explain any significant variation in the residuals. Taken together, these two results indicate that the chosen instrument will be appropriate.

2.5.2. Wage models

First, Table 2.6 shows the results of OLS estimations of Mincer's equations. Regarding the variables related to personal characteristics and job characteristics, the expected results are obtained in all three specifications. Being a woman and an immigrant are handicaps in terms of wage. Working in companies of 10 or fewer workers and living in areas with a low population density also result in lower wages. In contrast, greater potential work experience and a longer time working in a company have positive effects on wages. Finally, workers in high-skilled occupations with permanent contracts, working full time, earn higher wages than workers with few qualifications and temporary contracts working part time. Moreover, Heckman's lambda shows that self-selection of workers is relevant. The parameter is positive and statistically significant, so we cannot reject the possibility that there is a positive selection effect on wages, since it is probably the case that people with a greater probability of being employed earn more than the average for people who are employed and whose wages are reported.

Regarding the variables of interest in the study, all three equations show that the workers' years of schooling are statistically significant in the three specifications of the Mincer equation and have a positive return of approximately 3%. The results of equation [2] show that people who have undertaken non-formal education earn wages 4.7% (exp (0.0428) -1 = 0.0468) higher than those workers who have not done so, independent of their level of education. The results of equation [3] show that the returns on non-formal education are related to attained education, and its effect is greater if the worker has a higher level of education by just one year. Specifically, the return is 0.9% for each year of formal education.

Table 2.6. OLS estimates of the Mincer wage equations

	(1)	(2)	(3)
Woman	-0.206***	-0.207***	-0.208***
	(0.0103)	(0.0103)	(0.0103)
Immigrant	-0.0589***	-0.0541***	-0.0550***
	(0.0180)	(0.0179)	(0.0178)
Years of schooling	0.0322***	0.0313***	0.0284***
	(0.00202)	(0.00202)	(0.00227)
NFE		0.0428***	-0.0654*
		(0.0102)	(0.0366)
Years of schooling x NFE			0.00901***
			(0.00302)
Experience	0.0132***	0.0130***	0.0135***
	(0.00199)	(0.00198)	(0.00201)
Experience squared	-0.000209***	-0.000203***	-0.000215***
	(3.66e-05)	(3.65e-05)	(3.70e-05)
Seniority	0.00899***	0.00901***	0.00880***
	(0.00187)	(0.00186)	(0.00186)
Seniority squared	-5.99e-05	-6.18e-05	-5.43e-05
	(5.57e-05)	(5.55e-05)	(5.55e-05)
High skill occupation	0.237***	0.232***	0.227***
	(0.0146)	(0.0146)	(0.0146)
Permanent contract	0.0653***	0.0663***	0.0663***
	(0.0134)	(0.0134)	(0.0133)
Full-time job	0.443***	0.439***	0.439***
	(0.0213)	(0.0213)	(0.0212)
More than one job	-0.00600	-0.00580	-0.00410
	(0.0234)	(0.0233)	(0.0233)
Company size (<10 workers)	-0.0908***	-0.0872***	-0.0866***
	(0.0117)	(0.0116)	(0.0116)
Heckman's λ	0.000511***	0.000490***	0.000518***
	(0.000147)	(0.000135)	(0.000160)
Constant	6.006***	6.004***	6.034***
	(0.0435)	(0.0434)	(0.0435)
Observations	5180	5180	5180
R-squared	0.516	0.518	0.519

Note: NFE=non-formal education. Standard errors in parentheses. *** statistically significant at 1%; ** statistically significant at 5%; * statistically significant at 10%. All models include economic sector dummies and area density and regional dummies. Individual sample weights are considered.

A more complex picture emerges when we replace years of education with years of over-, required- and undereducation following the ORU specification, as the results do not hold for all workers.

Table 2.7 shows the results of the weighted ORU wage equation. Regarding the control variables, the results are similar to those found in Mincer's equations. As previously mentioned, we have used the statistical method based on the mode in order to calculate *Sr*, *So* and *Su*. Column 1 shows that both the years of education required for the job and the years of either overeducation or undereducation have a significant effect on wages. In particular, it can be seen that an additional required year of education has a return of 4.7% while an additional year of overeducation results in a return of 3.4% compared to the wage earned by workers in the same occupation whose actual education matches that required for the job; and, finally, the return of an additional year of undereducation is -2.3% compared to the wage earned by workers with the same level of education for whom job matches occupation. These findings are similar to those found in the literature (Hartog, 2000; Rubb, 2003; Verhaest and Omey, 2006). We can also see from this table that workers who have undertaken non-formal education earn higher wages than the rest, with a value of the coefficient associated with the dummy variable similar to the one found before.

Column 2 of Table 2.7 shows the results of the same ORU specification presented above, but taking into account the interactions between the different variables related to the number of schooling years, skill mismatch and participation in non-formal education. In this way, we analyse whether workers participating in non-formal education activities receive different returns. Again, the parameters associated with the years of education are statistically significant, although the magnitudes are slightly different in the case of required years and years of overeducation.

Table 2.7. OLS estimates of the ORU wage equations

	(1)	(2)
Woman	-0.208***	-0.209***
	(0.0103)	(0.0103)
Immigrant	-0.0510***	-0.0501***
	(0.0180)	(0.0180)
NFE	0.0408***	-0.0481
	(0.0101)	(0.0530)
Years of required education	0.0464***	0.0439***
	(0.00302)	(0.00346)
Years overeducation	0.0332***	0.0277***
	(0.00333)	(0.00419)
Years undereducation	-0.0237***	-0.0235***
	(0.00295)	(0.00343)
NFE x Years of required education		0.00642
		(0.00421)
NFE x years overeducation		0.0143**
		(0.00614)
NFE x years undereducation		-0.00179
		(0.00590)
Experience	0.0142***	0.0144***
	(0.00200)	(0.00202)
Experience squared	-0.000228***	-0.000235***
	(3.69e-05)	(3.75e-05)
Seniority	0.00863***	0.00851***
	(0.00186)	(0.00186)
Seniority squared	-5.71e-05	-5.22e-05
	(5.58e-05)	(5.59e-05)
High skill occupation	0.193***	0.190***
	(0.0155)	(0.0156)
Permanent contract	0.0627***	0.0629***
	(0.0133)	(0.0133)
Full-time job	0.439***	0.439***
	(0.0211)	(0.0211)
More than one job	-0.00515	-0.00336
	(0.0231)	(0.0232)
Company size (< 10 workers)	-0.0894***	-0.0890***
	(0.0116)	(0.0116)
Heckman's λ	0.000533***	0.000547***
	(0.000133)	(0.000148)
Constant	5.823***	5.854***
	(0.0513)	(0.0534)
Observations	5180	5180
R-squared	0.523	0.524

Note: Standard errors in parentheses. *** statistically significant at 1%; ** statistically significant at 5%; * statistically significant at 10%. All models include economic sector dummies and area density and regional dummies. Individual sample weights are considered.

Regarding the interactions between non-formal education and the different components of schooling years (over-, required- and undereducation) included in column 2, we can see that the coefficient associated with overeducation is positive and statistically significant. In other words, those overeducated workers who perform non-formal education activities receive a 1.4% higher salary for a year of overeducation than overeducated workers who do not perform those types of learning activities. Moreover, we do not find a statistically significant effect between the other interactions.

So, overeducated workers have a wage penalty compared to properly matched workers with the same level of education. However, overeducated workers who undertook non-formal education activities receive a higher wage than overeducated workers who do not perform those activities; then, they reduce the wage penalty they suffer to be in a job below their education level.

In conclusion, our results are in line with those of other studies, both international and for the case of Spain. In particular, it can be seen that the wage that an undereducated worker earns in a specific occupation tends to be less than that of workers with the level of studies required for that job, while the wage of an overeducated worker tends to be greater than that of properly educated workers in the same occupation, although less than they could expect to earn in an occupation that requires the higher level of education.

Moreover, workers who have engaged in non-formal education activities get a return on those activities, which is higher if the schooling level of the worker is higher. However, when the situation of educational mismatch is considered, we find that this wage premium is observed only for overeducated workers who perform non-formal education activities. Nevertheless, those wages would continue to be lower than those of workers with the same level of education in matched occupations. Although non-formal education does not increase people's level of schooling, it does contribute to improving the competences and skills of workers and, consequently, allows them to earn higher wages.

2.5.3. Robustness check

We perform the same analysis measuring educational mismatch by the mean version instead of the mode version in a robustness check of the results[9].

Table A2.3 in the Annex shows the results of the OLS estimation of the ORU equation. As shown in column (1), the effect of required years of schooling is positive and higher than the effect of years of overeducation. In contrast, the effect of years of undereducation is negative. When the interactions between educational mismatch variables and non-formal education are included, the results are the same of those obtained in column (2) of Table 2.7. In particular, overeducated workers who perform non-formal education activities receive a higher salary for a year of overeducation than overeducated workers who do not perform those types of learning activities. So, the two measures of educational mismatch point out the same conclusion, although, as Rubb (2003) finds, the magnitude of the coefficient varies.

2.6. Conclusions

Although investment in education at early ages is still a priority of most governments, important efforts are being devoted to promote lifelong learning since it permits both individuals and society to better adapt to changes in economic conditions.

The objective of this chapter was twofold. First, it examined the effect of non-formal education activities on individual wages and tested whether this effect is different depending on workers' years of schooling. And, second, it analysed whether the returns of the participation in these types of training activities is higher for overeducated workers than for the rest of workers. Using microdata from the Spanish sample of the 2007 Adult Education Survey, we have determined that non-formal education activities have a positive effect on

[9] Some additional robustness checks have been carried out in order to analyse the stability of the results with the inclusion/exclusion of different variables related to employment characteristics (such as activity sector or occupation) without any relevant change in terms of the conclusions obtained.

wages, this effect being higher for those workers with higher education. Moreover, we found that overeducated workers suffer a wage penalisation compared to well-educated workers with the same level of education. However, only overeducated workers who have undergone non-formal education activities receive a wage premium. It seems that this type of training provides overeducated workers with new abilities that permit them to reduce the wage penalisation derived from the mismatch between their level of education and occupation.

A tentative explanation of these results, which should be further explored in future research, is that non-formal education activities permit overeducated workers to obtain new abilities and better adjust to the requirements of their current jobs. This could be related to the fact that there are a number of companies who have difficulties filling their job vacancies with suitable candidates. Then, companies may decide to hire workers who have a higher level of education required for the job, but with a higher facility to get training to adjust their skills and education for the job. So, once they get the specific training, they obtain a wage return. From the point of view of public policy, our results suggest that lifelong learning activities should be promoted in those occupations with a higher incidence of overeducation, i.e., those that experience more trouble filling their vacancies with suitable candidates. However, although the situation has improved during recent years, nowadays, statistics on vacancies in European countries and, in particular, in Spain are not sufficiently developed to permit a rigorous test of our hypothesis.

References

Alba-Ramírez, A. (1993) "Mismatch in the Spanish Labor Market: Overeducation?", *The Journal of Human Resources*, 28, 259-278.

Aguilar, M. I. and García, D. (2008) "Desajuste educativo y salarios en España: nueva evidencia con datos de panel", *Estadística Española*, 50(168), 393-426.

Badillo-Amador, L., García-Sánchez. A. and Vila, L. E. (2005) "Mismatches in the Spanish labor market: Education vs. competence match", *International Advances in Economic Research*, 11(1), 93-109.

Barone, C. and Ortiz, L. (2011) "Overeducation among European University Graduates: a comparative analysis of its incidence and the importance of higher education differentiation", *Higher Education*, 61(3), 325-337.

Bauer, T. (2002) "Educational mismatch and wages: a panel analysis", *Economics of Education Review*, 21, 221–229.

Beneito, P.; Ferri, J.; Molto, M.L. and Uriel, E. (1996) "Desajuste educativo y formación laboral especializada: efectos sobre los rendimientos salariales", *IVIE, WP-EC 96-11*.

Blanden, J.; Buscha, F.; Sturgis, P. and Urwin, P. (2012) "Measuring the earnings returns to lifelong learning in the UK", *Economics of Education Review*, 31, 501-514.

Bound, J.; Jaeger, D.A. and Baker, R.M. (1995) "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistical Association*, 90(430), 443-50.

Büchel, F. (2002) "The effects of overeducation on productivity in Germany – the firms' viewpoint", *Economics of Education Review*, 21(3), 263-275.

Büchel, F. and Mertens, A. (2004) "Overeducation, undereducation, and the theory of career mobility", *Applied Economics*, 36, 803-816.

De Coulon, A. and Vignoles, A. (2008) "An Analysis of the Benefit of NVQ2 Qualifications Acquired at Age 26-34" DP106, *Centre for the Economics of Education,* LSE.

Dolton P.; Vignoles A. (2002) "Is a broader curriculum better?", *Economics of Education Review*, 21(5), 415-429.

Duncan, G. and Hoffman, S. (1981) "The incidence and wage effects of overeducation", *Economics of Education Review*, 1(1), 75-86.

Egerton, M. and Parry, G. (2001) "Lifelong Debt: Rates of Return to Lifelong Learning", *Higher Education Quarterly*, 55, 4-27.

Freeman, R. (1976) *The Overeducated American*. Academic Press.

García-Montalvo, J. (1995) "Empleo y sobrecualificación: el caso español", *FEDEA*, Working Paper 95-20.

García-Montalvo, J. (2005) "La inserció laboral dels universitaris: la qüestió de la sobrequalificació i el desajust formatiu", *Nota d'Economia*, 1, 99-119.

García Serrano, C. and Malo, M.A. (1996) "Desajuste Educativo y Mercados Internos de Trabajo en España", *Revista de Economía Aplicada*, 11(4), 105-131.

Groot, W. and Maassen van den Brink, H. (2000) "Overeducation in the Labor Market: a Meta-Analysis", *Economics of Education Review*, 19(2), 149-158.

Hartog, J. (2000) "Over-education and earnings: Where are we, where should we go?", *Economics of Education Review*, 19, 131-147.

Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-161.

Hersch, J. (1991) "Education match and job match", *Review of Economics and Statistics*, 73(1), 140-144.

Jenkins, A.; Vignoles, A., Wolf, A. and Galindo-Rueda, F. (2002) "The Determinants and Effects of Lifelong Learning", *Applied Economics*, 35, 1711-1721.

Kiker, B.F.; Santos, M.C. and Mendes de Oliveira, M. (1997) "Overeducation and Undereducation: Evidence for Portugal", *Economics of Education Review*, 16(2), 111-125.

Leuven, E. and Oosterbeek, H. (2011) "Overeducation and Mismatch in the Labor Market", in: E. Hanushek, S. Machin, L. Woessman (Eds.), *Handbook of the Economics of Education*, 4, 283-326.

Mas, M.; Perez, F.; Uriel, E.; Serrano, L. (1995) *Capital Humano Series Históricas, 1964-1992*, Fundación Bancaixa.

Mincer, J. (1974) *Schooling, Experience and Earnings*. Columbia University Press, New York.

Murillo, I.P.; Rahona, M. and Salinas, M. (2012) "Effects of educational mismatch on private returns to education: An analysis of the Spanish case (1995-2006)", *Journal of Policy Modeling*, 34(5), 646-659.

OECD (2007) *International Migration Outlook*, Paris.

OECD (2009) *Education at a glance 2009*, Paris.

Oliver, J.; Raymond, J.L., (dirs.) (2002) *Educación formal y demanda de cualificación de la mano de obra en España. Una visión a largo plazo.* Centre d'Economia Industrial, Document d'Economia Industrial n. 14, 136 p.

Quintini, G. (2011) "Over-Qualified or Under-Skilled: A Review of Existing Literature*", OECD Social, Employment and Migration Working Papers*, No. 121, OECD Publishing.

Ramos, R, and Sanromá, E (2013) "Overeducation and local labour markets in Spain", *Tijdschrift voor Economische en Sociale Geografie*, 104(3), 278-291.

Rosen, S. (1976) "A theory of life earnings", *Journal of Political Economy*, 84(4), S45-S67.

Rubb, S. (2003) "Overeducation in the labor market: A comment and re-analysis of a meta-analysis", *Economics of Education Review*, 22(6), 621–629.

Sloane, P. (2003) "Much ado about nothing? What does the overeducation literature really tell us" In Büchel, de Grip and Mertens (eds), *Overeducation in Europe* (pp.11–49). Cheltenham: Edward Elgar.

Verdugo, R. and Verdugo, N. (1989) "The impact of surplus schooling on earnings", *Journal of Human Resources*, 24(4), 629-643.

Verhaest, D. and Omey, E. (2006) "The impact of overeducation and its measurement", *Social Indicators Research*, 77(3), 419-448.

Van Smoorenburg, M. and van der Velden, R. (2000) "The training of school-leavers: Complementarity or substitution?", *Economics of Education Review*, 19(2), 207-217.

Annex chapter 2

Table A2.1. Equivalent level of studies and schooling years

Higher level of schooling	Schooling years
No schooling / illiterates	0
Primary education	6
First stage of secondary education without school graduate degree or equivalent	8
First stage of secondary education with school graduate degree or equivalent	10
Vocational-technical education of short duration	10
Middle level of vocational-technical education	12
Undergraduate education	12
Vocational-technical education of short duration or middle level, or undergraduate education	12
Higher level of vocational-technical education	12
University 1st cycle and equivalent, or 3 full graduate courses or equivalent credits	15
University 1st and 2nd cycle, only 2nd cycle and equivalent	17
Official studies of professional expertise	17
Postgraduate university education	19

Table A2.2. Descriptive statistics

Variable	Sample 1 Mean	Sample 1 Std. Dev.	Sample 2 Mean	Sample 2 Std. Dev.	Min	Max
Logarithm monthly wage	6.99	0.44	7.00	0.44	4.13	9.10
Woman	0.41	0.49	0.42	0.49	0	1
Immigrant	0.13	0.33	0.15	0.35	0	1
Schooling years	11.47	3.61	11.50	3.54	0	19
Non-formal education	0.32	0.47	0.38	0.48	0	1
Potential experience	23.48	11.60	22.30	11.18	2	62
Experience squared	685.85	624.17	622.57	583.25	4	3844
Seniority	9.76	9.84	8.92	9.35	0	49
Seniority squared	192.05	336.92	166.82	307.64	0	2401
Agriculture	0.05	0.23	0.04	0.20	0	1
Industry	0.17	0.37	0.19	0.39	0	1
Construction	0.11	0.32	0.11	0.32	0	1
Services	0.40	0.49	0.37	0.48	0	1
No sale services	0.27	0.44	0.28	0.45	0	1
Qualified occupation	0.29	0.45	0.25	0.43	0	1
Permanent contract	0.76	0.43	0.74	0.44	0	1
Full-time job	0.90	0.30	0.88	0.32	0	1
More than one job	0.05	0.23	0.05	0.23	0	1
Firm with 10 workers or less	0.32	0.47	0.29	0.45	0	1
High density population area	0.52	0.50	0.50	0.50	0	1
Low density population area	0.23	0.42	0.26	0.44	0	1
Qualified occupation	0.25	0.43	0.24	0.43	0	1
Balearic Islands	0.15	0.36	0.14	0.35	0	1
Canary Islands	0.03	0.17	0.02	0.15	0	1
Cantabria	0.02	0.15	0.02	0.14	0	1
Castilla León	0.03	0.16	0.03	0.17	0	1
Castilla La Mancha	0.05	0.21	0.05	0.21	0	1
Catalonia	0.01	0.11	0.02	0.13	0	1
Valencia	0.05	0.23	0.05	0.22	0	1
Extremadura	0.04	0.20	0.04	0.20	0	1
Galicia	0.17	0.38	0.16	0.37	0	1
Madrid	0.11	0.31	0.13	0.33	0	1
Murcia	0.02	0.14	0.03	0.16	0	1
Navarra	0.06	0.24	0.07	0.26	0	1
Basque Country	0.15	0.36	0.12	0.32	0	1
Rioja	0.03	0.17	0.05	0.21	0	1
Ceuta and Melilla	0.01	0.12	0.01	0.11	0	1
Balearic Islands	0.05	0.22	0.05	0.22	0	1
Canary Islands	0.01	0.08	0.01	0.10	0	1

Note: Sample 1 considers all workers; and Sample 2 restricts to workers with no missing values in all variables considered (5133 observations).

Table A2.3. OLS estimates of the ORU wage equations using the statistical method (mean)

	(1)	(2)
Woman	-0.219***	-0.219***
	(0.0106)	(0.0105)
Immigrant	-0.0363**	-0.0358**
	(0.0182)	(0.0183)
NFE	0.0433***	-0.00141
	(0.0102)	(0.0504)
Years of required education	0.0464***	0.0451***
	(0.00393)	(0.00434)
Years overeducation	0.0414***	0.0317***
	(0.00668)	(0.00818)
Years undereducation	-0.0397***	-0.0376***
	(0.00622)	(0.00711)
NFE x Years of required education		0.00329
		(0.00433)
NFE x years overeducation		0.0276**
		(0.0129)
NFE x years undereducation		-0.00960
		(0.0129)
Experience	0.0123***	0.0126***
	(0.00201)	(0.00203)
Experience squared	-0.000217***	-0.000223***
	(3.69e-05)	(3.72e-05)
Seniority	0.00873***	0.00866***
	(0.00188)	(0.00189)
Seniority squared	-5.44e-05	-5.18e-05
	(5.63e-05)	(5.64e-05)
High skill occupation	0.167***	0.166***
	(0.0198)	(0.0199)
Permanent contract	0.0702***	0.0698***
	(0.0134)	(0.0133)
Full-time job	0.433***	0.433***
	(0.0214)	(0.0214)
More than one job	-0.00139	-6.91e-05
	(0.0234)	(0.0235)
Company size (less than 10 workers)	-0.0906***	-0.0904***
	(0.0116)	(0.0116)
Lamda Heckman	0.000515***	0.000532***
	(0.000147)	(0.000164)
Constant	5.889***	5.902***
	(0.0541)	(0.0571)
Observations	5,180	5,180
R-squared	0.514	0.515

Note: Standard errors in parentheses. *** statistically significant at 1%; ** statistically significant at 5%; * statistically significant at 10%. All models include economic sector dummies and area density and regional dummies. Individual sample weights are considered.

Chapter 3: Overeducation, skills and wage penalty

3.1. Introduction

There is a remarkable consensus on the effects of educational mismatch on wages using the standard ORU specification (Duncan and Hoffman, 1981). On the one hand, undereducated workers benefit from a wage premium compared to well-educated workers with the same level of education. On the other hand, overeducated workers earn more than their properly educated co-workers, but earn less than they would at a job requiring their level of education. So, while undereducated workers earn more than their properly matched counterparts, overeducated workers experience a wage penalty.

One of the proposed theories to explain overeducation's wage penalty is based on the assignment theory (Sattinger, 1993). It considers that workers' productivity is limited by their workplace and in part by their human capital. So, overeducated workers may thus underutilize their skills, and, in consequence, they are less productive and obtain lower wages than well-educated workers with the same level of education. Following that idea, overeducation may imply overskilling. However, empirical evidence shows a weak correlation between both variables, which means that the assignment theory does not seem to be supported by data (Allen and van der Velden, 2001; Green and McIntosh, 2007).

A supported alternative theory is based on the existence of individuals' skill heterogeneity. From such a perspective, the wage penalty associated to overeducation is due to the huge variation of skills between workers with the same level of education. Then, overeducated workers would not suffer a wage penalty. In fact, they would earn lower wages as a result of their lower skills. If this hypothesis holds, the wage penalty will disappear once individuals' skill level is included in the analysis. However, most of the literature does not explicitly test this hypothesis due to data limitations regarding individuals' skill levels.

In this chapter we take advantage of the recently available database of the OECD Programme for the International Assessment of Adult Competencies (PIAAC) because it includes information about individual skills from proficiency test's scores. It allows testing whether individuals' skill heterogeneity could explain the effects of educational mismatch on wages.

We focus on the Spanish case because it has some interesting features that justify the analysis. As mentioned in previous chapter, it is a developed country supporting one of the largest percentages of overeducated workers (OECD, 2013a), a feature that was also observed before the current economic crisis (OECD, 2009; and Verhaest and van der Velden, 2013). At the same time, the Spanish population[10] has one of the lowest levels of proficiency in literacy and numeracy skills (OECD, 2013a).

Therefore, the specific aims of the chapter are twofold:

1) Test whether the assignment theory is supported or not by the Spanish data. With this aim we will perform a statistical analysis of the correlation between both educational and skill mismatches.

2) Test the individuals' skill heterogeneity theory in Spain. Our hypothesis is that the wage penalty associated to overeducation could be explained by their lower skill levels. In consequence, overeducated workers may not be suffering a wage penalty in Spain, but their earnings are determined by their skill level.

[10] Along with Italy (OECD, 2013a)

Our results show a weak correlation between educational and skill mismatches, as it is found in other analyses. Thus, the assignment theory does not seem to be supported by Spanish data. We also find that individuals' skill heterogeneity only explains 18% of the effect of educational mismatch on wages in Spain. The wage penalty still remains for those overeducated workers who are not less skilled than properly matched workers.

The rest of the chapter is structured as follows. First, section 3.2 provides a literature review on the analysis of skills in educational mismatch. Section 3.3 introduces the PIAAC data and explains how educational and skills mismatch are measured. Section 3.4 shows the relationship between overeducation and overskilling. Section 3.5 quantifies the wage penalty of overeducation and the impact of skills using different specifications. Section 3.6 concludes with some final remarks.

3.2. Literature review

Different theories have been considered in order to explain the overeducation phenomenon (see Leuven and Oosberbeek, 2011 and Quintini, 2011 for a review). However, the most frequently regarded are the assignment model and individuals' skills heterogeneity.

The assignment theory (Sattinger, 1993) makes the assumption that human capital returns depend on both the workers' human capital and the match between the worker and the job. The basic idea is that, although education raises productivity in general, working in a job below one's own qualification level imposes a ceiling on a worker's productivity because it limits the extent to which his or her skills can be utilised and results in lower wages. According to assignment theory, productivity is maximised when workers are allocated top-down according to their skills, whereby the most skilled are assigned to the most complex job and the least skilled to the simplest job. Overeducation is explained by differences in the share of complex jobs and skilled workers. So, overeducated workers may underutilize their skills and, in consequence, they are less productive and obtain lower wages than well-educated workers with the same level of education. Following that idea, overeducation may imply overskilling – or broadly speaking, educational mismatch may imply skill mismatch.

Thanks to the availability of recent databases providing questions relative to skill mismatch, the assignment theory has been explicitly tested. Skill mismatch has been measured by means of subjective workers' responses about whether they consider that their skills are used enough in their jobs. Following the specification developed by Verdugo and Verdugo (1989), different studies have included dummy variables for both educational and skill mismatch in the empirical analysis (Allen and van der Velden, 2001; Di Pietro and Urwin, 2006; Green and McIntosh, 2007; Sánchez-Sánchez and McGuinness, 2013; Mavromaras et al. 2013). It has been found that overeducation and overskilling have both a negative and statistically significant effect on earnings within the same level of education, the overeducation effect being much higher than the overskilling effect. This result underlines that wage penalization associated with overeducation is not explained by under-utilization or waste of workers' skills, whereas the assignment theory is not supported by the results. They may suggest the existence of heterogeneity of workers' skills. However, they do not explicitly test this theory due to a lack of information about workers' skill level rather than skill mismatches.

Specifically, the heterogeneous skills theory takes into account human capital differences between workers. It considers that workers' productivity depends on the human capital level acquired, regardless of job characteristics. Therefore, the observed wage differences among overeducated and undereducated workers compared to well-matched workers with the same educational level may only reflect individual differences in human capital within educational levels. In other words, overeducated workers may be less productive because they have less human capital, not because their job imposes limitations on their productivity.

As has been mentioned before, data availability on workers' skill levels is very limited, whereas different approaches have been considered in empirical analysis to attempt to control for individual skill heterogeneity in the wage equation estimation.

One approach involves the consideration of panel data sets in order to control for all unobserved individual fixed effects (Bauer, 2002; Frenette, 2004; Korpi and Tåhlin, 2009; Tsai, 2010). They find that the wage penalty associated with being overeducated falls dramatically and even disappears when it is estimated by fixed effects, suggesting that (part of) the effect of educational mismatch is caused by unobserved individual ability. Carroll and Tani (2013) find that unobserved heterogeneity explains the effect of overeducation on earnings only for young graduates.

Instead of using a longitudinal framework, Chevalier (2003) analyses cross-sectional data. In order to proxy the unobservable skills, he estimates earnings in the first job and assumed that the deviation between the expected and observed earnings is a proxy for the unobservable idiosyncratic characteristics affecting workers' productivity. In this case, after accounting for the unobserved heterogeneity, the wage penalty for overeducation is slightly reduced. Using a similar methodology, Chevalier and Lindley (2009) arrive at analogous results. They construct a measure of unobserved ability as the residual from a first-job earnings equation, capturing all individual's observed characteristics including job characteristics that affect wages. These residuals should then be a proxy for all time-invariant unobservable characteristics. Chevalier (2003) also introduces a new approach overlapping overeducation and workers' job satisfaction. He divides overeducation into two categories: 'apparent' overeducation, composed of satisfied graduate workers; and 'genuine' overeducation, consisting of dissatisfied graduate workers. Results show that 'genuine' overeducation brings a much larger pay penalty than 'apparent' overeducation.

Following this approach, Green and Zhu (2010) find similar results. They also consider different types of overeducation but use a direct measure of skills. Pecoraro (2014) also estimates a wage model considering 'genuine' and 'apparent' overeducation. He applies two estimation methods to control for unobserved ability: the standard fixed effects model and an OLS model that includes a proxy of unobserved ability as suggested by Chevalier (2003). He finds that the wage penalty is no more significant for graduates who are overeducated but well matched in skills.

On the other hand, the recent study of Levels et al. (2013) includes individuals' skill level in the analysis using PIAAC data. They analyse the effect of workers' skills level on the effect of educational mismatches derived from ORU specification for different OECD countries. They find that a considerable part of the effect of educational mismatches on wages can be attributed to skill heterogeneity, but it still remains statistically significant.

In summary, empirical evidence does not seem to support the assignment theory, given that there is a weak relation between educational and skill mismatches. The individuals' skill heterogeneity theory seem to be the most supported explanation for the observed wage differences between overeducated and properly matched workers.

Although there is a wide literature analysing the impact of overeducation on wages in Spain (see Alba-Ramírez, 1993; Murillo et al. 2012; among others), to our knowledge, no extensive analysis has tested the role of individuals' skill level on educational mismatch focusing on the Spanish case.

3.3. Data sources and variable definition

3.3.1. PIAAC database

The Programme for the International Assessment of Adult Competencies (PIAAC) is a survey which has been conducted by the OECD. It assesses the proficiency of adults from age 16 onwards in literacy, numeracy and problem solving in technology-rich environments. In addition, the survey collects a range of information on reading, writing and numeracy-related activities of respondents, as well as education, labour and family background variables. It was conducted in 24 countries (22 OECD countries) between 2011 and 2012.

Participation in the problem-solving domain was optional, and Spain (and other countries) did not participate in it. As a consequence, the competences we analyse are related to literacy and numeracy. Specifically, the two domains are defined in the following way:

- Literacy: ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential. Literacy encompasses a range of skills from the decoding of written words and sentences to the comprehension, interpretation, and evaluation of complex texts.

- Numeracy: ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. To this end, numeracy involves managing a situation or solving a problem in a real context, by responding to mathematical content/ information/ideas represented in multiple ways.

Both literacy and numeracy are measured by 10 plausible values calculated using Item Response Theory (IRT), which are represented on a 500-point scale. The idea is that each individual only responds to a limited number of items in the test. To avoid the assignation of missing values in those items which have not been included in the test, the procedure predicts scores using answers from the test and background questionnaires of similar individuals. It generates a distribution of values for each individual and their associated probabilities, with ten plausible values randomly obtained for each individual. This method prevents bias from estimating the result from a small number of test questions. We also consider the jackknife method (80 replicate weights) implemented in PIAAC to derive standard errors in wage regressions[11].

Given the high correlation between literacy and numeracy skill level (0.92), we only perform the next analysis using literacy skills. However, we repeat the whole analysis using numeracy skills instead of literacy skills as a robustness check.

We consider two sets of variables. The first one includes variables related to workers' human capital as years of education (derived from levels of education), experience, experience squared, non-formal education, and 10

[11] See OECD (2013b): Technical Report of the Survey of Adult Skills (PIAAC) for more details about IRT and the Jackknife method.

plausible values test scores in literacy. The second one is composed of other personal, job related and regional variables that are included in the model as controls[12]. These variables are gender, age, nationality, type of contract (full-time/part-time), contact term (temporary/permanent), sector (public/private), economic activity (industry, agriculture, construction, services, non-sale services) and 17 regions.

The initial Spanish sample was composed of 6055 observations. We restrict the sample to employed workers who were not enrolled in education at the time. We drop from the analysis armed forces workers, and participants who did not give some of the information we need to perform the analysis. The final sample was 1928 observations. Table A3.1 of the Annex shows the descriptive analysis of the variables previously defined.

3.3.2. Measuring educational and skill mismatches

As it has been explained in section 1.2 of chapter 1, there are different methods to measure educational mismatch. The PIAAC data allows us to measure required schooling using both the worker's self-assessment and the statistical method.

The self-assessment method relies on questions that ask workers about the schooling requirements of their job. The PIAAC questionnaire specifically contains the following questions: *"If applying today, what would be the usual qualifications, if any, that someone would need to get this type of job?"*. Educational mismatch is obtained by comparing workers' answers about required education and attained education. Workers are properly or well-matched when their attained education matches with their jobs' required education. Conversely, overeducated (undereducated) workers have more (less) attained education than required by their jobs.

[12] The estimation results for these explanatory variables will not be discussed. A full set of the estimation results is available on request.

The statistical method (both mean and mode versions) uses information about workers' schooling and their occupations. Regarding the mean version, the required amount of schooling for a worker is determined by the mean of attained education of all workers holding the same occupation. Workers are then defined to be overeducated or undereducated if their attained education deviates at least one standard deviation from the mean in their occupation. The mode version measure required schooling from the mode of attained education of all workers holding the same occupation. It classifies overeducated or undereducated workers according to whether their education differs from the mode in their occupation.

Table 3.1 shows the impact of educational mismatch in Spain using the self-assessment method. All the following results have been obtained using the individual sample weights. About half of workers in Spain have a proper match between their education and occupation. From the remaining workers, the PIAAC data highlights that overeducation is affecting 35.63%, 3.8 being the average number of surplus years of education. On the other hand, undereducation concerns the other 15% and their average number of deficit years of education is 3.1.[13] [14]

Table 3.1. Educational mismatch

	Percentage	Average mismatch in years of education
Undereducation	15.17	3.10 (deficit)
Proper education	49.20	0.00
Overeducation	35.63	3.80 (surplus)

The percentages obtained by the statistical method are shown in Table A3.2 of Annex of this chapter. It is worth noting that different measurement methods were used to report different percentages of educational mismatch, although they are considered to be the same country and the same database. However, the impact on wages is consistent regardless of the measurement method

[13] Although OECD (2013a) measures educational mismatch using the same self-assessment method than us, the percentages of mismatch are different. The reason of those differences is that they cluster education into 4 levels while we take advantage of the maximum level of disaggregation of the data.
[14] Similar incidence of educational mismatch in Spain has been found by Murillo et.al. (2012)

considered (Hartog, 2000). We perform the main analysis measuring educational mismatch using the self-assessment method, and we repeat it using the statistical method as a robustness check.

With regards to the measurement of skill mismatch, we follow the approach defined by the OECD using PIAAC data (Pellizzari and Fichen, 2013; OECD, 2013a). It is a combination of workers' self-assessment questions and their skill proficiency score. The survey asks workers whether they feel they "*have the skills to cope with more demanding duties than those they are required to perform in their current job*" and whether they feel they "*need further training in order to cope well with their present duties*". To compute the OECD measure of skills mismatch, workers are classified as well-skilled in a domain if their skill proficiency score in that domain is between the minimum and maximum score observed among workers who answered "no" to both questions in the same 1-digit occupation (and country). Workers are over-skilled in a domain if their score is higher than the maximum score of self-reported well-skilled workers, and they are under-skilled in a domain if their score is lower than the minimum score of self-reported well-skilled workers.

Individual weighted results show that 72% of workers have a good match between their skills and those required by their jobs. Moreover, overskilling affects 21.4% of workers whereas 6.5% are underskilled.

3.4. Are overeducated workers also overskilled?

As it has been explained in the previous sections, individuals' skill heterogeneity is one of the explanations of the fact that assignment theory does not seem to be supported by empirical evidence. In other words, most studies have usually found a weak correlation between overeducation and overskilling.

In this section, we analyse the correlation between both educational and skill mismatch (Table 3.2) to check whether the assignment theory is supported or not using data for Spain. We also compare the distribution of skills between different types of workers to find differences that could suggest the existence of individuals' skill heterogeneity.

The PIAAC data for Spain shows that all workers have a higher probability of being well-skilled, regardless of their education-occupation (mis)match. In particular, we find that 72% of undereducated and 70% of overeducated workers are well-skilled in their jobs. It is surprising that only 7.5% of undereducated workers are also underskilled and 20% have an excess of skills. However, the data shows that 23% of overeducated workers are also overskilled. This results is consistent with Allen and van der Velden (2001) and Green and McIntosh (2007). Indeed, the Pearson chi-square test formally validates the lack of correlation between educational and skill mismatch in Spain[15].

Table 3.2. Distribution of undereducated, well-matched and overeducated workers by their skill (mis)match in literacy (in %)

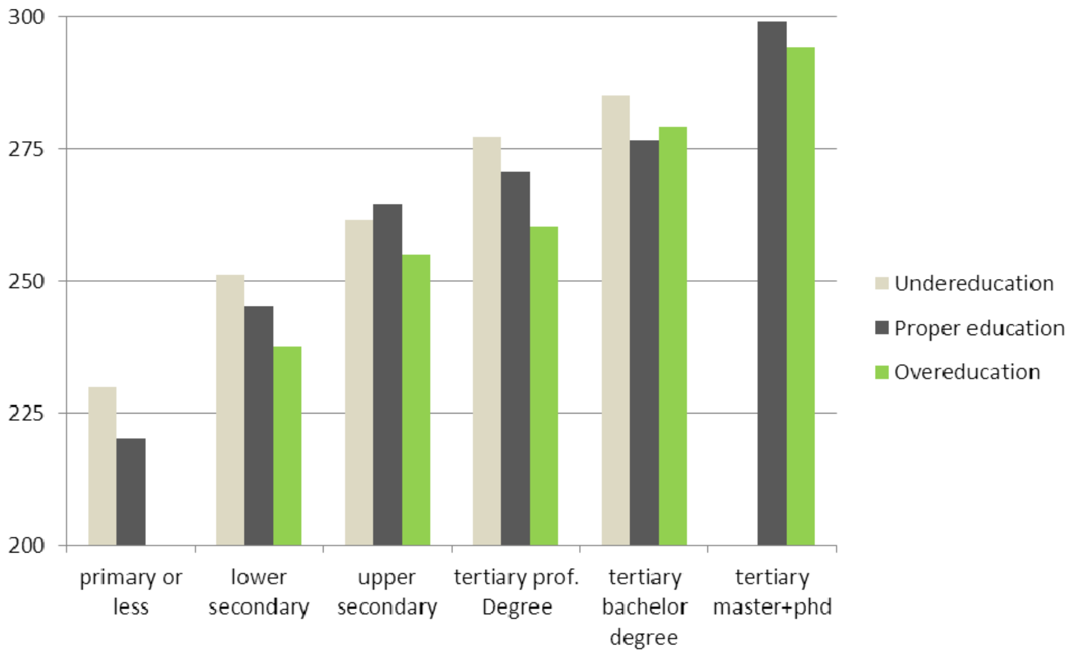
	Underskilling	Proper skills	Overskilling	Total
Undereducation	7.48	72.24	20.28	100.00
Proper education	6.52	73.18	20.29	100.00
Overeducation	6.02	70.62	23.36	100.00

Thus, the empirical evidence for Spain does not seem to support the assignment theory, since educational mismatches are not associated to skill mismatches.

Figure 3.1 shows the skill level of workers by educational mismatches and by different levels of education in order to provide preliminary evidence existence of skill heterogeneity between workers. It shows that overeducated workers hold a lower skill level than properly educated workers with the same educational level. That fact is repeated for all educational levels (except for bachelor degree). However, undereducated workers tend to have a higher skill level than properly-educated workers with the same educational level (except for upper secondary education).

[15] The Pearson chi-square test rejects the null hypothesis of non-correlation between variables. Pearson chi2(4) = 4.1182 Pr = 0.390.

Figure 3.1. Average skills levels of workers by educational level.



Thus, the data show skill heterogeneity between workers with the same level of education. This fact could explain the wage differences between workers according to their education-occupation match. This is empirically tested in the following section 3.5.

3.5. Educational mismatch, skills and wages

3.5.1. Empirical models

In order to quantify the effect of educational mismatch on wages, different specifications based on the traditional wage equation (Mincer, 1974) have been proposed in the literature: the ORU specification developed by Duncan and Hoffman (1981) and the Verdugo and Verdugo (1989) specification. The traditional wage model considers formal education as a proxy of individuals' human capital. However, it is well known that there are components of human capital such as skills or ability. We therefore also include individuals' skills in all three models.

Specifically, the traditional wage equation is defined as follows:

$$\log(W_i) = \alpha + \beta \cdot S_i^a + \theta^{'} \cdot X_i + u_i \qquad (1a)$$

where $\log(W_i)$ is the logarithm of the hourly wage of worker i; S_i^a refers to the number of years of formal education; X_i is a vector of control variables related to personal, job and regional characteristics that also includes other human capital variables such as experience, experience squared and a dummy variable that takes the value of 1 if the worker has participated in some non-formal education activity during the last 12 months prior to the survey and 0 otherwise. Finally, u_i is the error term with zero mean and constant variance.

Including the individuals' proficiency skills ($skills_i$), the modified model is then defined as:

$$\log(W_i) = \alpha + \beta \cdot S_i^a + \gamma \cdot skills_i + \theta^{'} \cdot X_i + u_i \qquad (1b)$$

where $skills_i$ is a continuous variable measured by scores in a 500-point scale. The higher the score is, the higher the individual's skill level.

A variant of the traditional Mincerian wage equation is the ORU (Over-Required-Under-educated) specification created by Duncan and Hoffman (1981). This specification splits years of education (S^a) into three variables: years of education required for the job (S^r), years of overeducation (S^o) and years of undereducation (S^u).

Specifically, it holds that $S^a = S^r + S^o - S^u$

In this sense, the following is determined:

- $S^o = S^a - S^r$ if the worker is overeducated and 0 if otherwise, and
- $S^u = S^r - S^a$ if the worker is undereducated and 0 if otherwise.

The ORU equation is then defined as:

$$\log(W_i) = \alpha + \beta_1 \cdot S_i^r + \beta_2 \cdot S_i^o + \beta_3 \cdot S_i^u + \theta' \cdot X_i + u_i \qquad (2a)$$

The other variables' definitions are the same as in specification (1a). The interpretation of the coefficients associated with over- and undereducation is compared with well-matched workers in the same job. The usual findings in the literature are $\beta_1 > \beta_2 > |\beta_3|$.

In order to test the individual's skill heterogeneity hypothesis, we also include the variable related to individuals' skills:

$$\log(W_i) = \alpha + \beta_1 \cdot S_i^r + \beta_2 \cdot S_i^o + \beta_3 \cdot S_i^u + \gamma \cdot skills_i + \theta' \cdot X_i + u_i \quad (2b)$$

The variable $skills_i$ is defined as in equation (1b). If individuals' skills heterogeneity completely explains the wages' effects of educational mismatch, we should get $\beta_2 = \beta_3 = 0$. If this is true, workers' remuneration composed by their education and skills would be determined by the required education and their individual skill level.

Another contribution to the overeducation literature has been defined by Verdugo and Verdugo (1989, henceforth V&V). This model includes dummy variables related to overeducation and undereducation using the Mincerian wage equation.

The V&V equation is defined as:

$$\log(W_i) = \sigma_0 + \sigma_1 \cdot S_i^a + \sigma_2 \cdot OE_i + \sigma_3 \cdot UE_i + \theta' \cdot X_i + u_i \quad (3a)$$

where OE is a dummy variable that takes the value 1 when the worker is overeducated and 0 otherwise, and UE is also a dummy variable that takes the value 1 when the worker is undereducated and 0 otherwise. The coefficients associated with both variables show the average wage effect of being overeducated and undereducated compared with well-matched workers with the same level of education. The usual finding is that overeducated workers have a wage penalization and undereducated workers benefit from a wage premium compared to well-matched workers with the same educational level. That is, $\sigma_2 < 0$ and $\sigma_3 > 0$.

We also extend that model including skill level variable. The extended V&V model is then defined as follows:

$$\log(W_i) = \sigma_0 + \sigma_1 \cdot S_i^a + \sigma_2 \cdot OE_i + \sigma_3 \cdot UE_i + \rho_1 \cdot skills_i^a + \theta^{'} \cdot X_i + u_i \qquad (3b)$$

In the case that the individual's skills heterogeneity theory is valid, we expect that both coefficients associated with overeducation and undereducation are not statistically significant once we control for individuals' skills. If this is so, workers would be remunerated by their attained education and skills level.

3.5.2. Results

In line with similar studies (see, for instance, Dolton and Vignoles, 2000; and Di Pietro and Urwin, 2006), we control in all the previous specifications for a possible problem of sample selection bias estimating using Heckman two step specification (Heckman, 1979). This procedure takes into account the possibility that employed workers may not be a random subsample of the sample we are considering. The first step estimates the probability of being employed using a probit equation[16] (see the results in Table A3.3). Then, the probit estimation is used to construct a selection bias control factor, which is included as an explanatory variable in the wage equation[17].

As we focus on the analysis of the variables related to human capital, we only comment on the results of those variables in the main test. However, it is worth noting that the coefficients associated to control variables are similar to those in the previous literature. Furthermore, it is found that the lambda coefficient is positive and statistically significant for all specifications. Hence,

[16] The probit equation of the probability of being employed includes as explanatory variables gender, experience, experience squared, years of attained education, immigrant status, number of children, whether individual is living with spouse or not, and regional dummies.
[17] The variables we use as exclusion restrictions are both number of children at home and whether individual is living with partner or not. Those variables affect the probability of being employed, but do not determine wages.

the omission of the information about the probability of being employed in the wage analysis would imply a bias in the results.

Table 3.3 reports the results from the estimations of the Mincerian wage models specified in equations (1a) and (1b), the ORU models defined in equations (2a) and (2b) and the V&V specifications defined in equations (3a) and (3b).

With respect to the traditional Mincer's models, it is shown that the returns of the variables related to human capital are similar to previous literature findings (column 1). The return of attained education is 6.4% per year. The years of experience in work also has a positive impact on wages, but there is a moment that its positive impact is decreasing. Finally, training activities in non-formal education also has a positive and significant effect on wages (13%).

When individuals' skills are included in the model (column 2), we find a positive statistically significant effect on wages. Specifically, for each skill's score, individuals have a return of 0.14%. The magnitude of the effect of skills may seem small, but it is important to remember that skills are measured by scores in a 500-point scale. Furthermore, the coefficients of the other variables related to human capital (education, experience and non-formal education) are reduced once skills are included.

Regarding the ORU specification defined in equation (2a), we find that the return of required education is higher than the return of attained education. It points out the existence of educational mismatch. Contrary of most of literature, we find that the return of one year of overeducation is lower than the return (in absolute term) of one year of undereducation. In particular, overeducated workers obtain for each surplus year of education a 3% higher salary than well-educated workers in the same job. Undereducated workers obtain a 3.7% lower wage than well-educated workers in the same job.

In order to test the individuals' skill heterogeneity theory, individuals' skills are included explicitly in the ORU model as specified in equation (2b). Skills have a positive and statistically significant effect on wages, but the effects of educational mismatch still remain statistically significant. Thus, for each year of required education, wage increases 7.45%. The coefficient related with years of overeducation falls from 3% to 2.4% and the coefficient for years of undereducation decreases from 3.7% to 3.1%. Indeed, the hypothesis that years of overeducation and years of undereducation are both equal to 0 (ie. $\beta_2 = \beta_3 = 0$) after controlling for skill is rejected at a level of 1% significance. Hence, the obtained results show that individual's skills heterogeneity explains only part of the wage effects of educational mismatch, and therefore our initial hypothesis about the Spanish case is not supported by the data. Specifically, skills only explain 18% of the wage's effect of overeducation and 14% of the effect of undereducation on wages. The obtained results are in line with the analysis of Levels et al. (2013) for a set of OECD countries.

Finally, the results from V&V specifications defined in equations (3a) and (3b) are shown in the last columns of table 3.4. The effects of both dummy variables related to overeducation and undereducation are in line with previous literature. Overeducated workers suffer a wage penalization compared to well-educated workers with the same level of education while undereducated workers earn higher wages than well-educated workers with the same level of education. Once individual's skills are included in the model (equation 3b), the effects of educational mismatches are very slightly reduced. Specifically, the penalty associated with overeducated workers is reduced from 17.1% to 16.3%. On the other hand, the premium of undereducated workers falls from 13.1% to 12.4%. As found in the results from the ORU specifications, these results do not seem to support the heterogeneity skills theory, since the effects of overeducation and undereducation still remain once skill is controlled for.

To sum up, we find that individuals' skills are important to determine individuals' wages as well as other human capital variables. However, contrary to our initial hypothesis, we do not find that individuals' skills heterogeneity completely explains the effect of educational mismatch on wages. Specifically, the lower skills of overeducated workers only explain 18% of their lower wages compared to well-matched workers with the same level of education.

Table 3.3. Estimated earnings functions

	Mincer		ORU		V&V	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Male	0.208***	0.189***	0.193***	0.179***	0.195***	0.180***
	[0.0224]	[0.0186]	[0.0218]	[0.0185]	[0.0219]	[0.0178]
Immigrant	-0.108***	-0.0728**	-0.0728**	-0.0471	-0.0840**	-0.0550
	[0.0357]	[0.0355]	[0.0350]	[0.0339]	[0.0355]	[0.0347]
Attained education (years)	0.0617***	0.0529***			0.0706***	0.0627***
	[0.00509]	[0.00482]			[0.00504]	[0.00494]
Required education (years)			0.0715***	0.0642***		
			[0.00530]	[0.00517]		
Overeducation (years)			0.0295***	0.0243***		
			[0.00657]	[0.00597]		
Undereducation (years)			-0.0361***	-0.0310***		
			[0.00919]	[0.00831]		
Overeducation (dummy)					-0.158***	-0.151***
					[0.0211]	[0.0207]
Undereducation (dummy)					0.123***	0.117***
					[0.0309]	[0.0253]
Skill level (scores)		0.00144***		0.00111***		0.00122***
		[0.000291]		[0.000293]		[0.000292]
Experience	0.0170***	0.0165***	0.0172***	0.0168***	0.0170***	0.0166***
	[0.00393]	[0.00389]	[0.00389]	[0.00383]	[0.00381]	[0.00385]
Experience2	-0.000206**	-0.000173**	-0.000246***	-0.000218***	-0.000236***	-0.000207**
	[8.40e-05]	[8.45e-05]	[8.20e-05]	[8.12e-05]	[8.15e-05]	[8.19e-05]
Non-formal education	0.128***	0.120***	0.0974***	0.0934***	0.103***	0.0984***
	[0.0220]	[0.0204]	[0.0213]	[0.0198]	[0.0214]	[0.0202]
Full-time	-0.0379	-0.0307	-0.0512	-0.0449	-0.0520	-0.0452
	[0.0352]	[0.0350]	[0.0350]	[0.0359]	[0.0348]	[0.0358]
Permanent	0.106***	0.0995***	0.0907***	0.0869***	0.0896***	0.0852***
	[0.0265]	[0.0280]	[0.0261]	[0.0276]	[0.0260]	[0.0272]
Public sector	0.135***	0.137***	0.151***	0.151***	0.150***	0.150***
	[0.0355]	[0.0376]	[0.0337]	[0.0355]	[0.0338]	[0.0349]
Activity sector	Yes	Yes	Yes	Yes	Yes	Yes
Regions	Yes	Yes	Yes	Yes	Yes	Yes
Lambda mills	0.115**	0.116***	0.0871*	0.0896**	0.0968**	0.0986***
	[0.0512]	[0.0396]	[0.0526]	[0.0406]	[0.0487]	[0.0377]
Constant	0.677***	0.442***	0.692***	0.511***	0.681***	0.483***
	[0.149]	[0.123]	[0.150]	[0.125]	[0.143]	[0.120]
Observations	1928	1928	1928	1928	1928	1928
R-squared	0.392	0.404	0.432	0.438	0.425	0.432
H_o: $\beta_2 = \beta_3 = 0$				23.35***		

Standard errors in parentheses. *Statistically significant at the 10% level. **Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered. Equations (1b), (2b) and (3b) also take into account the 10 plausible values of skill level and the 80 replications weights in both estimations.

3.5.3. Robustness checks

The PIAAC data allows us to perform some robustness checks to validate the previous results.

First, literature shows that the incidence of both overeducation and undereducation could be different depending on the measurement method applied. However, the effects on wages are quite consistent regardless of the measurement method. Besides the self-assessment method, the PIAAC data allows us to measure educational mismatch by means of both versions of the statistical method, the mean and the mode. The results from the ORU specification measuring educational mismatch by means of both statistical methods confirm the main results (Table A3.4. of Annex). Specifically, individuals' skills only explain 14% of the wage penalty of overeducated workers in both models.

Second, we use the variable skill level in numeracy instead of skill level of literacy. As we have already notice, both variables are highly correlated, and therefore we decided not to include both together. We also estimate the ORU specifications including numeracy skills instead of literacy skills (Table A3.5. of Annex). The results show that the wage penalty of overeducated workers is reduced but still remains once skills are included. Specifically, individual's skills heterogeneity explains 22% of the wage penalty.

3.6. Final remarks

The main objective of this paper is to analyse whether individual's skill heterogeneity explains the wage penalty of being overeducated in Spain. Our hypothesis is that the wage penalty associated with overeducation could be explained by the low skill level of overeducated workers, since Spain holds the lowest level of skill among its population. As a consequence, overeducated workers may not be suffering a wage penalty in Spain, and otherwise their earnings are determined by their skill level. Our results show that individuals' skill heterogeneity only explains 18% of the effect of educational mismatch on wages in Spain. The wage penalty of overeducated workers still remains for those who are not less skilled than properly educated workers.

There are some policy recommendations associated with the previous results. On the one hand, as part of the effect of overeducation on wages is due to a lack of competence or skills of overeducated workers, educational policy makers should focus on defining the level of competence or skills that should be acquired at each level of education. Indeed, skills should be evaluated at educational institutions in the same way as education.

On the other hand, other measures should be taken into account by policy makers in Spain, since the wage penalty still remains after controlling for individual skills. First, educational institutions should give all the information about the employability of each type of education to students before they start a specialized course. Second, they should also encourage students in entrepreneurship. Self-employment could be a way to overcome the lack of demand for specific workers. Finally, the Spanish government should make an effort to promote the creation of companies that require high-skilled workers to create a production system based on high technologies.

References

Alba-Ramírez, A. (1993) "Mismatch in the Spanish labor market: overeducation?", *The Journal of Human Resources*, 27, 259–78.

Allen, J. and van der Velden, R. (2001) "Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search", *Oxford Economic Papers*, 53, 434–452.

Bauer, T. (2002) "Educational Mismatch and Wages: A Panel Analysis." *Economics of Education Review*, 21, 221–229.

Carroll, D. and Tani, M. (2013) "Over-education of recent higher education graduates: New Australian panel evidence", *Economics of Education Review*, 32, 207-218.

Chevalier, A. (2003) "Measuring Overeducation", *Economica*, 70 (3), 509-531.

Chevalier, A. and Lindley, J. (2009) "Overeducation and the skills of UK graduates", *Journal of the Royal Statistical Society: Series A*, 172, 307-337.

Di Pietro, G. and Urwin, P. (2006) "Education and skills mismatch in the Italian graduate labour market", *Applied Economics*, 38(1), 79-93

Dolton, P. and Vignoles, A. (2000) "The incidence and the effects of overeducation in the UK graduate labour market", *Economics of Education Review*, 19, 179–98.

Duncan, G. and Hoffman, S. (1981) "The incidence and wage effects of overeducation", *Economics of Education Review*, 1(1), 75-86.

Frenette, M. (2004) "The overqualified Canadian graduate: the role of the academic program in the incidence, persistence, and economic returns to overqualification", *Economics of Education Review*, 23, 29-45.

Green, F. and McIntosh, S. (2007) "Is There a Genuine Under-utilisation of Skills Amongst the Over-qualified?", *Applied Economics*, 39(4), 427-439.

Green, F. and Zhu, Y. (2010) "Overqualification, Job Dissatisfaction and Increasing Dispersion in the Returns to Graduate Education", *Oxford Economic Papers,* 62(4), 740-763.

Hartog, J. (2000) "Overeducation and earnings: where are we, where should we go?", *Economics of Education Review*, 19, 131–147.

Heckman, J.J. (1979) "Sample selection bias as a specification error", *Econometrica*, 47, 153–61.

Korpi, T. and Tåhlin, M. (2009) "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000", *Labour Economics*, 16, 183-193.

Leuven, E. and Oosterbeek, H. (2011) "Overeducation and Mismatch in the Labor Market", in: E. Hanushek, S. Machin, L. Woessman (Eds.), *Handbook of the Economics of Education*, 4, 283-326.

Levels, M., van der Velden, R. and Allen, J. (2013) "Educational mismatches and skills: new empirical tests of old hypotheses", *RM*/13/062.

Mavromaras, K.; McGuinness, S.; O'Leary, N.; Sloane, P. and Wei, Z. (2013) "Job Mismatches and Labour Market Outcomes: Panel Evidence on University Graduates", *Economic Record*, 89, 382-395.

Mincer, J. (1974) *Schooling, Experience, and Earnings*, New York: NBER Press.

Murillo, I.P.; Rahona-López, M.; Salinas-Jiménez. MdM. (2012) "Effects of educational mismatch on private returns to education: An analysis of the Spanish case (1995–2006)", *Journal of Policy Modeling*, 34(5), 646-659.

OECD (2009) *Education at a glance 2009*, Paris.

OECD (2013a) *OECD Skills Outlook 2013 First Results from the Survey of Adult Skills*, Paris.

OECD (2013b) *Technical Report of the Survey of Adult Skills (PIAAC),* Paris.

Pecoraro, M. (2014) "Is there still a wage penalty for being overeducated but well-matched in skills?" A panel data analysis of a Swiss graduate cohort" *Labour*, forthcoming.

Pellizzari, M. and Fichen, A. (2013) "A New Measure of Skills Mismatch: Theory and Evidence from the Survey of Adult Skills (PIAAC)", *OECD Social, Employment and Migration Working Papers*, 153, OECD Publishing.

Quintini, G. (2011) "Over-Qualified or Under-Skilled: A Review of Existing Literature", *OECD Social, Employment and Migration Working Papers*, No. 121, OECD Publishing.

Sánchez-Sánchez, N. and McGuinness, S. (2013) "Decomposing the impacts of overeducation and overskilling on earnings and job satisfaction: an analysis using REFLEX data", *Education Economics*, forthcoming.

Sattinger, M. (1993) "Assignment models of the distribution of earnings", *Journal of Economic Literature*, XXXI, 831-880.

Tsai, Y. (2010) "Returns to overeducation: A longitudinal analysis of the U.S. labor market", *Economics of Education Review,* 29, 606–617.

Verdugo, R. and Verdugo, N. (1989) "The impact of surplus schooling on earnings: some additional findings", *Journal of Human Resources*, 24 (4), 629-643.

Verhaest, D. and van der Velden, R. (2013) "Cross-country Differences in Graduate Overeducation", *European Sociological Review*, 29(3), 642-653.

Annex chapter 3

Table A3.1. Descriptive statistics

	Mean	Std. Dev.	Min	Max
Log(wage)	2.18	0.50	0.18	4.53
Literacy proficiency	260.07	42.87	78.76	367.19
Numeracy proficiency	256.71	44.50	82.32	380.86
Age	41.32	10.08	16	65
Male	0.55	0.50	0	1
Immigrant	0.13	0.33	0	1
Attained education	12.21	3.43	6	21
Experience	18.45	10.66	0	55
Experience squared	453.88	470.44	0	3025
Non-formal education	0.56	0.50	0	1
Full time job	0.85	0.36	0	1
Permanent contract	0.81	0.39	0	1
Public sector	0.25	0.43	0	1
Agriculture	0.04	0.19	0	1
Construct	0.07	0.25	0	1
Services	0.47	0.50	0	1
No-sale services	0.28	0.45	0	1

Note: Number of observations 1928.

Table A3.2. Educational mismatch using the statistical method

	Mean	Std. Dev	Min	Max
MODE				
Overeducation	0.3516185	0.4775994	0	1
Proper education	0.3975564	0.4895198	0	1
Undereducation	0.2508252	0.4336005	0	1
Years overeducation	3.3781	1.771671	1	11
Years undereducation	3.599314	1.629411	1	11
MEAN				
Overeducation	0.1409579	0.3480684	0	1
Proper education	0.6873661	0.4636868	0	1
Undereducation	0.171676	0.3771964	0	1
Years overeducation	1.30805	1.07916	0.0111046	4.856499
Years undereducation	1.429313	1.103022	0.1687933	6.62323

Table A3.3. Determinants of being employed

	(1)
Male	0.202***
	[0.0458]
Experience	0.0702***
	[0.00676]
Experience2	-0.00109***
	[0.000155]
Attained education	0.106***
	[0.00649]
Immigrant	-0.0564
	[0.0673]
Number of children	-0.0674***
	[0.0247]
Living with spouse	0.0158
	[0.0557]
Regional dummies	Yes
Constant	-1.996***
	[0.145]
Observations	4689

Note: Estimated coefficients. Standard errors in parentheses. *Statistically significant at the 10% level. **Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered.

Table A3.4. Estimated earnings ORU functions measuring educational mismatch by means of the statistical method (mode and mean)

	Mode		Mean	
	(1a)	(1b)	(2a)	(2b)
Male	0.202***	0.187***	0.209***	0.193***
	[0.0216]	[0.0181]	[0.0212]	[0.0181]
Immigrant	-0.0765**	-0.0488	-0.0626*	-0.0362
	[0.0352]	[0.0348]	[0.0350]	[0.0336]
Required education (years)	0.0823***	0.0741***	0.0977***	0.0884***
	[0.00550]	[0.00547]	[0.00602]	[0.00630]
Overeducation (years)	0.0431***	0.0369***	0.0720***	0.0618***
	[0.00662]	[0.00657]	[0.0168]	[0.0158]
Undereducation (years)	-0.0393***	-0.0330***	-0.0799***	-0.0690***
	[0.00767]	[0.00714]	[0.0141]	[0.0142]
Skill level (scores)		0.00118***		0.00123***
		[0.000284]		[0.000281]
Experience	0.0161***	0.0157***	0.0148***	0.0147***
	[0.00382]	[0.00384]	[0.00340]	[0.00363]
Experience2	-0.000214***	-0.000187**	-0.000211***	-0.000186**
	[8.17e-05]	[8.40e-05]	[7.77e-05]	[8.16e-05]
Non-formal education	0.102***	0.0976***	0.0978***	0.0928***
	[0.0220]	[0.0208]	[0.0224]	[0.0212]
Full-time	-0.0365	-0.0307	-0.0397	-0.0334
	[0.0340]	[0.0350]	[0.0342]	[0.0360]
Permanent	0.101***	0.0964***	0.105***	0.0985***
	[0.0255]	[0.0273]	[0.0252]	[0.0255]
Public sector	0.130***	0.131***	0.142***	0.141***
	[0.0348]	[0.0362]	[0.0350]	[0.0365]
Activity sector	Yes	Yes	Yes	Yes
Regions	Yes	Yes	Yes	Yes
Lambda	0.0793	0.0819**	0.0496	0.0610**
	[0.0511]	[0.0400]	[0.0306]	[0.0264]
Constant	0.492***	0.309**	0.379***	0.183*
	[0.147]	[0.120]	[0.123]	[0.111]
Observations	1928	1928	1928	1928
R-squared	0.429	0.437	0.429	0.438

Standard errors in parentheses. *Statistically significant at the 10% level. **Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered. Column s (1b) and (2b) also take into account the 10 plausible values of skill level and the 80 replications weights in both estimations.

Table A3.5. Estimated earnings ORU functions using numeracy skills

	(1a)	(1b)
Male	0.193***	0.169***
	[0.0218]	[0.0190]
Immigrant	-0.0728**	-0.0424
	[0.0350]	[0.0336]
Required education (years)	0.0715***	0.0624***
	[0.00530]	[0.00539]
Overeducation (years)	0.0295***	0.0231***
	[0.00657]	[0.00603]
Undereducation (years)	-0.0361***	-0.0310***
	[0.00919]	[0.00830]
Skill level (scores)		0.00131***
		[0.000299]
Experience	0.0172***	0.0163***
	[0.00389]	[0.00386]
Experience2	-0.000246***	-0.000205**
	[8.20e-05]	[8.19e-05]
Non-formal education	0.0974***	0.0906***
	[0.0213]	[0.0195]
Full-time	-0.0512	-0.0420
	[0.0350]	[0.0360]
Permanent	0.0907***	0.0840***
	[0.0261]	[0.0277]
Public sector	0.151***	0.150***
	[0.0337]	[0.0354]
Activity sector	Yes	Yes
Regions	Yes	Yes
Lambda	0.0871*	0.0908**
	[0.0526]	[0.0412]
Constant	0.692***	0.497***
	[0.150]	[0.123]
Observations	1928	1928
R-squared	0.432	0.441

Note: Standard errors in parentheses. *Statistically significant at the 10% level. **Statistically significant at the 5% level. ***Statistically significant at the 1% level. Individual sample weights considered. Column (1b) also takes into account the 10 plausible values of skill level and the 80 replications weights in both estimations.

Chapter 4: Is there a link between parents' overeducation and children's educational achievement? [18]

4.1. Introduction

The analysis of the determinants of students' performance at school has been the focus of researchers during the last few decades. It is found that variables related to students' characteristics, as well as family and school characteristics, have a significant impact on students' educational achievement. However, the set of variables comprised of the students' family characteristics, especially parents' human capital, seems to be among the most important in order to explain students' educational achievement (Coleman et al., 1966).

To the best of our knowledge, one aspect that has not been considered so far in the group of variables related to parents is the possible relationship between parents' educational mismatch and children's educational achievement.

As Haveman and Wolfe (1995) highlight, children of highly educated parents tend to perform better than children of less educated parents. In this sense, one of the arguments to explain the better performance of students whose parents have a high educational level is the genetic transmission of cognitive

[18] Previous versions of Chapter 4 have been presented in XXI Jornadas de la Asociación de Economía de la Educación, 2012 (Oporto, Portugal), and XV Encuentro de Economía Aplicada, 2012 (A Coruña, Spain). One of the previous versions has been published as "Nieto, S and Ramos, R. (2011) ¿La sobreeducación de los pedres afecta al rendimiento educativo de sus hijos?, Regional and Sectoral Economic Studies / Estudios Económicos Regionales y Sectoriales, 11(3), 97-118.

skills (see Björklund and Salvanes, 2010, for a review). In particular, Plug and Vijverberg (2003) quantify that about half of the intergenerational transmission of human capital is genetically transmitted. However, there are other possible explanations for the intergenerational effect of human capital besides genetic transmission. One possible explanation for the positive relationship between parents' human capital and students' performance is based on children's perceptions about the importance of education. In this sense, students whose parents have a high level of education and good jobs might be more aware of the value of education and, consequently, have a higher motivation and perform better than other students.

Under this point of view, our hypothesis is that the existence of parents' job-education mismatch can modify the students' perception about the importance of education and, consequently, have an effect on their performance at school. Moreover, this effect may be different depending on whether parents are overeducated or undereducated. For instance, children whose parents are overeducated – i.e., those with a higher educational level than their job requires – may have a "discouragement" effect in school derived from the experience of their parents. Specifically, they may have the perception that the opportunity cost of investing in education will exceed future profits obtained by increasing their levels of human capital. In other words, they may underestimate the benefits derived from education due to the bad match of their parents' education in the labour market. In this case, we would expect lower educational achievement for those students whose parents are overeducated in relation to students whose parents are properly educated. As for the perception of children whose parents are undereducated, the result is less clear. On the one hand, students may have the perception that there is no need to study to get a good job, because their low-educated parents have secured relatively good jobs. On the other hand, students could also be aware of the troubles that their parents might suffer as a result of training deficiencies for being undereducated. Then, students might put more effort into their educational outcomes. Therefore, for this second group it is more difficult to identify, a priori, the final effect of parents' undereducation on students' educational achievement.

So, taking the previous information into account, the questions we empirically analyse are the following:

- Is there a relationship between parents' educational mismatch and the educational performance of their children?

- In case this relationship exists, is this effect similar across the performance distribution or, by contrast, are there differences between students at the top and at the bottom of the performance distribution?

We carry out the analysis for Spain because it has some peculiarities that are not found in other developed countries. Firstly, Spain has one of the highest percentages of overeducation among the OECD countries: around 33% of the population (Quintini, 2011). The quantification of the overeducation phenomenon in a country is important given its negative effects on wages and job satisfaction of overeducated workers (see Hartog, 2000; and Leuven and Oosterbeek, 2011; for a review). Then, the higher the percentage of the overeducated population, the higher the percentage of the population suffering from lower wages and job dissatisfaction in the country.

And, secondly, educational outcomes of students in Spain are constantly being challenged whenever PISA results are published, since students' educational outcomes in Spain are below the OECD average in all subjects considered: Mathematics, Sciences and Reading. In order to improve the students' outcomes, the Spanish government has already carried out three educational reforms[19] since the first publication of PISA results in 2000. However, after those reforms, PISA data show that the average performance of students in Spain remains below the average of the OECD countries and the differences between good and bad students have increased over the years (OECD, 2010).

Using microdata from the PISA 2009 wave, our results show a statistically significant relationship between parents' educational mismatch and children's educational performance after controlling for the effect of individual and

[19] LOCE (Ley Orgánica de Calidad de la Educación) promoted by the Popular Party in 2002, the LOE (Ley Orgánica de Educación) promoted by the Social Party in 2006 and the LOMCE (Ley Orgánica para la mejora de la calidad educativa) promoted by the Popular Party in 2013.

school characteristics and other family-related variables. On the one hand, students whose parents are overeducated have a penalty in their academic achievement in all three subjects analysed, this effect being stronger for students with lower educational outcomes. On the other hand, undereducation only affects students' educational achievement when it is suffered by the mother, this effect being positive. So, the results confirm our hypothesis, although they cannot prove that the students' perception is the transmission channel, an aspect that is beyond the scope of the current research due to data limitations.

The remainder of the chapter is organised as follows. Section 4.2 presents a literature review of the determinants of educational achievement. Section 4.3 describes the database and the variables used in the empirical analysis, including measures of parents' educational mismatch. Then, section 4.4 describes the methodological approach used, and the results obtained are shown. Finally, we summarise the main conclusions.

4.2. Literature Review

Different studies have identified many determinants of student's educational achievement, which, by their nature, can be grouped into three groups. The first group is composed of students' individual characteristics, such as gender, age, student's country of origin and student's native language. For instance, Marks (2008) shows that boys tend to outperform girls in mathematics while results are opposite in reading. Regarding students' country of origin, it is found that immigrants have lower educational achievement than natives (Meunier, 2011; Ammermueller, 2007; Entorf and Lauk, 2008; Calero and Waisgrais, 2008; for Spain). There is also evidence of the differences in educational achievement depending on the language spoken in students' homes. In this sense, Entorf and Minoui (2005) show that immigrants who speak the official language in their home environment improve their academic performance.

The second group of variables is related to students' family background. Coleman et al. (1966) were the pioneers in showing the impact of family-related variables on students' educational achievement. It is found that the higher the education level of parents, the higher the educational achievement of their children (Feinstein and Symons, 1999; Häkkinen et al., 2003; and Woessmann, 2003). In addition, the family's economic level also has a positive effect on students' educational achievement (Dahl and Lochner, 2012; and Perelman and Santin, 2011; for Spain). On the other hand, there are differences in the effect of family structure on students' achievement. In particular, students with single parents tend to perform worse in school than students living with both parents (Hampden-Thompson, 2013). Finally, the cultural environment in students' homes also has an impact on students' educational achievement. As a consequence of data limitations, some studies use the number of books available in the students' home as a proxy of their cultural background. Several studies find a positive relationship between the cultural level in students' homes and their educational achievement (Meunier, 2011; Martins and Veiga, 2010).

Finally, the third group of variables is related to school characteristics, such as type and size of school, number of students per teacher student and peer effects. As has been explained, most of the analyses agree on the influence of individual and family characteristics on students' educational achievement, but there is no consensus among researchers regarding the influence of the variables relating to schools on students' achievement. Studies such as those by Coleman and Hoffer (1987), Hanushek (1986), Stevans and Sessions (2000) and Opdenakker and Van Damme (2006), among others, find that students' educational achievement is better in private schools than in public schools. However, other studies such as those by Noell (1982), Sander (1996) and Smith and Naylor (2005) show a null effect of ownership of the school on students' performance. Using data from the PISA 2003 wave and multilevel analysis, Calero and Escardíbul (2007) examine the effect of schools' ownership on students' educational performance in the Spanish educational system. They find that differences in the students' scores observed in favour of private schools are not due to the ownership of the school. The explanation is that students in private schools come from a high socio-economic environment, so they are in a better educational climate (with fewer immigrants), which contributes to improving their educational performance. On the other hand, the effect of school size on student performance is also unclear. Barnett et al. (2002) find a positive relationship between school size and educational achievement of students, while Hanushek and Luque (2003) show no significant effect of this variable in most of the analysed countries. Also, there are inconclusive results regarding the number of students per teacher. Krueger (2003) shows that students give a better performance in small classes, while Hanushek (2003) and Rivkin et al. (2005) find no statistically significant effect of this variable on students' educational performance. In contrast, there is a consensus on the importance of peers on students' educational achievement (see, for instance, Feinstein and Symons, 1999). Although the inclusion of the school variables in the analysis allows identifying differences in students' outcomes derived from school characteristics, we do not carry out a detailed analysis of these factors. In particular, given that the aim of the paper is not focused on the detailed effect of school variables on students' achievement, we include school fixed effects in the analysis, as this will allow us to control results via the heterogeneity among students derived from schools' differences.

4.3. Database and variables

4.3.1. PISA database

The database we use in our analysis is the Programme for International Student Assessment (PISA) coordinated by the OECD. The aim of PISA is to assess students' education when they reach the final stage of compulsory education at 15 years in 3 educational domains: Mathematics, Sciences and Reading literacy. PISA defines them in the following way:

- Reading literacy: understanding, using, reflecting on and engaging with written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

- Mathematical literacy: individuals' capacity to formulate, employ and interpret mathematics in a variety of contexts. This includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. Mathematical literacy also helps individuals recognise the role that mathematics plays in the world and make the well-founded judgements and decisions needed by constructive, engaged and reflective citizens.

- Scientific literacy: individuals' scientific knowledge, and use of that knowledge, to identify questions, acquire new knowledge, explain scientific phenomena and draw evidence-based conclusions about science-related issues; their understanding of the characteristic features of science as a form of human knowledge and enquiry; their awareness of how science and technology shape our material, intellectual and cultural environments; and their willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen.

In addition, the PISA database provides information about students' personal characteristics and other characteristics related to their family background and schools. The survey started in the year 2000, and, since then, it has been conducted every three years. Although Spain has participated in PISA from the beginning of the survey, the amount of information has been growing along the waves. So, we employ data from 2009[20]. The initial full Spanish sample is of 25,887 students and is representative at the national level across the regions of Andalusia, Aragón, Astúrias, the Balearic Islands, the Canary Islands, Cantabria, Castilla y León, Catalonia, Galicia, La Rioja, Madrid, Murcia, Navarre, Basque Country, Ceuta and Melilla.[21]

Regarding the variables of students' performance, the PISA database measures them by five plausible values for each subject. The plausible values are not the actual test scores. They are random numbers from the distribution of results that may be reasonably assigned to each individual. This methodology was developed by Mislevy and Sheehan (1987, 1989) and is based on the theory of imputation of missing values developed by Rubin (1987). It considers that each individual answers a limited number of items. In order to avoid the assignation of missing values in the questions not included in the given questionnaire, the values for those items are estimated. The results are

[20] Now PISA 2012 is available. However, this was not available when the present research was conducted. It would be interesting to perform the same analysis to check the possible different effect of the economic crisis.
21 PISA 2009 microdata is available at the following link: http://pisa2009.acer.edu.au/

predicted using the answers of similar individuals considering the context questionnaires. Instead of predicting a single score, it generates a distribution of values for each individual and their associated probabilities to randomly obtain five plausible values for each individual. This method prevents the bias from estimating the result from a small number of test questions. Plausible values contain components of random error variance, which are not optimal as individual test scores. Thus, this procedure is not suitable for the diagnosis of individuals, but only to estimate consistent population parameters. The plausible values are considered as dependent variables in our model.

4.3.2. Variables and final sample

The variables we consider in the analysis are the following. As for individual characteristics, we consider gender, age, students' nationality (native, first and second immigrant generation[22]) and whether the language of the PISA test is the same as the main language spoken in the student's home. Regarding the variables related to the student's family, we include family structure (nuclear, single parent and mixed), parents' educational level (measured as years of education[23]) and their International Socioeconomic Index of Occupational Status (ISEI). The ISEI is considered as a proxy of household income, since this information is not available in the PISA database for Spain. The ISEI is derived from student responses on parental occupation using the method proposed by Ganzeboom et al. (1992). In particular, this index is built from the occupational classification and takes values from 16 to 90. The lower the index's value, the lower the socioeconomic status, whereas the higher the index's value, the higher the socioeconomic status. Furthermore, the student's cultural background variable is approximated by the number of books available in the student's home. As we are not interested in the effect of school variables on students' performance and the PISA database identifies students'

[22] Native students are those students born in Spain, or those with at least one parent born in Spain; students who were born abroad with at least one parent born in Spain are also classified as 'native' students). First-generation students are those born outside Spain and whose parents were also born in another country, whereas second-generation students are those born in Spain but whose parents were born in another country.

[23] The PISA database provides data about parents' educational level following the ISCED classification. We have created a new variable transforming ISCED levels to years of education by means of the correspondence defined in the PISA report.

schools (in Spain there are 889 schools), we include school fixed effects instead of a set of variables related to school characteristics. This allows us to control results via the heterogeneity among students derived from schools' differences in all estimations. And we also avoid the loss of a significant number of observations from the missing values of most of school's variables. Last, taking into account that some educational competences are transferred to the Spanish regions (Spanish Autonomous Communities), regional fixed effects are also included in the model.

The final sample is restricted to students whose parents[24] were both working part-time or full-time at the time of the survey. We do not consider students who do not provide the necessary information to measure parents' educational mismatch such as parents' educational level and occupation. So, the final sample is composed of 13.098 observations. A summary of the variables we use in the analysis is presented in Table A-1 of the Annex.

4.3.3. Measuring educational mismatch

The variables of interest in the study are related to parents' job-education mismatch or educational mismatch. There are different methods to measure educational mismatch: the objective, the subjective or workers' self-assessment and the statistical or realised matches. All of them have advantages and disadvantages, whereas using either measure method depends on the availability of the data (see Hartog, 2000; and Leuven and Oosterbeek, 2011; for a review). Given the available information of the PISA database (both parents' education and parents' occupation[25]), it is only possible to measure educational mismatch by the statistical method.

The statistical method (both mean and mode versions) determines whether workers are overeducated, undereducated or properly educated, comparing workers' attained education and required education by their occupations. Regarding the mean version (Verdugo and Verdugo, 1989), the required

[24] The PISA database considers parents as those who spend more time with students. Then, a parent could be considered as a father or step-/foster father or male guardian, and mother or step-/foster mother or female guardian.
[25] We use a 3-digit disaggregation level.

amount of schooling for a worker is determined by the mean of attained education of all workers holding the same occupation. Workers are then defined as being overeducated or undereducated if their attained education deviates at least one standard deviation from the mean in their occupation. The mode version (Kiker et al. 1997) measures required schooling from the mode of attained education of all workers holding the same occupation. It classifies overeducated or undereducated workers according to whether their education differs from the mode in their occupation.

The percentages of educational mismatch using the mode and the mean version for fathers and mothers are shown in Table 4.1. All the results have been obtained using the individual sample weights.

Table 4.1. Parents' educational mismatch (in percentages)

	Mode		Mean	
	Fathers	Mothers	Fathers	Mothers
Undereducation	32.15	32.55	16.88	16.47
Proper education	43.17	47.71	68.35	71.02
Overeducation	24.68	19.74	14.77	12.51
Total	100	100	100	100

The table shows that the percentages of educational mismatch differ by the measurement method applied. On the one hand, the statistical method based on the mode shows percentages of both overeducation and undereducation higher than those obtained using the method based on the mean. Those differences are usual findings in the literature, but its effects on wages are consistent regardless of the measurement method considered (Hartog, 2000). These results are similar to those obtained in the Chapter 1 of the thesis and other studies for the Spanish case.

Moreover, Table 1 shows differences in the incidence of overeducation between fathers and mothers, whereas the incidence of undereducation is quite similar between both groups. Specifically, the percentage of overeducated fathers is 24.68% using the mode method (14.77% using the mean), while the percentage of mothers who are overeducated is around 20%. Given that the mean version has more limitations than the mode version

(Hartog, 2000)[26], we perform the main analysis measuring educational mismatch through the mode version. However, the results have been replicated measuring educational mismatch by the statistical method based on the mean as a robustness check in section 5.3.

4.4. Descriptive analysis

In order to show some descriptive evidence, we first show the density functions of students' educational performance in Mathematics, Sciences and Reading according to their parents' educational mismatch (Figures 4.1–4.3). This analysis allows us to identify whether there are different patterns of students' outcomes depending on their parents' educational mismatch. As we can see, density functions of students whose mothers are overeducated are located further to the left of the graphs, where the students' educational outcome is lower. The functions related to students whose mothers are properly educated or undereducated both show a similar distribution, although it is slightly lower in the case of students whose mothers are undereducated. In contrast, the density functions of students' educational outcome are very similar for those whose fathers are either overeducated or undereducated. Both are located in the bottom of the distribution in comparison to the students whose fathers are properly educated.

In summary, these graphs show that, without considering the effect of other factors, the educational performance of students whose parents have an educational mismatch appears to be penalised compared to students whose parents do not have educational mismatch. However, it is necessary to perform further analysis to determine whether students' educational achievement gap is associated with the educational mismatch of their parents.

[26] Hartog (2000) explains that the mean version shows similar percentages for both overeducation and undereducation as a consequence of the symmetry of the normal distribution tails.

Figure 4.1. Students' educational outcomes distribution in Mathematics according to parents' educational mismatch

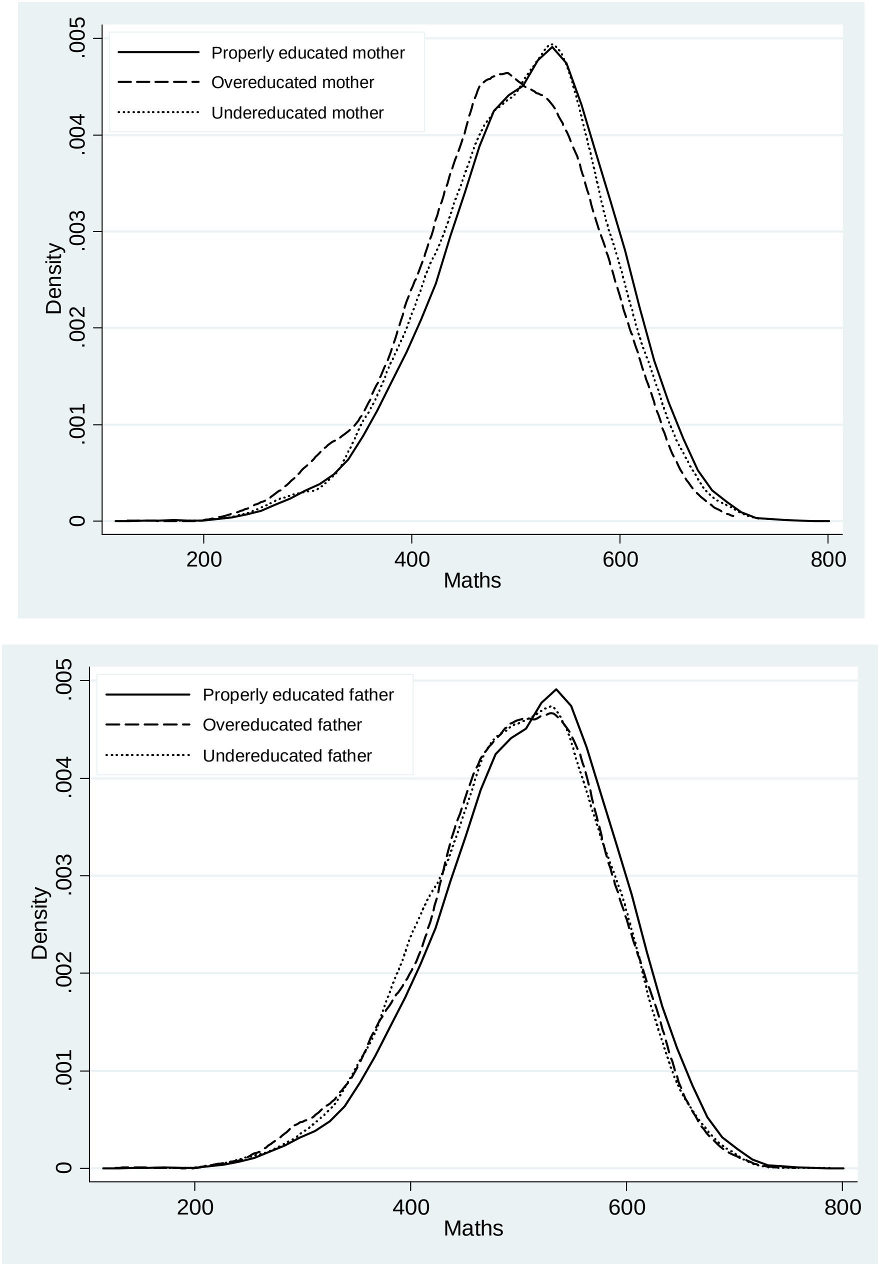
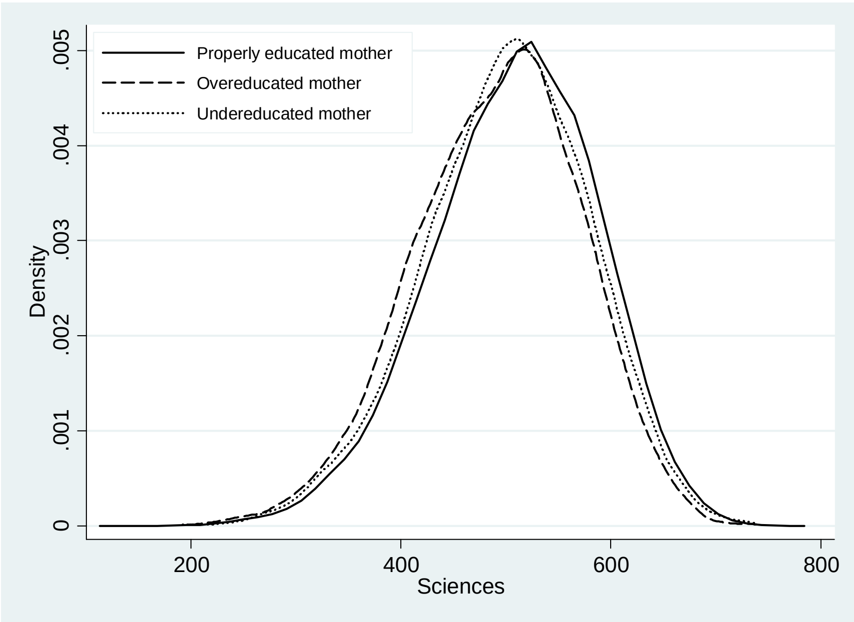


Figure 4.2. Students' educational outcomes distribution in Science according to parents' educational mismatch



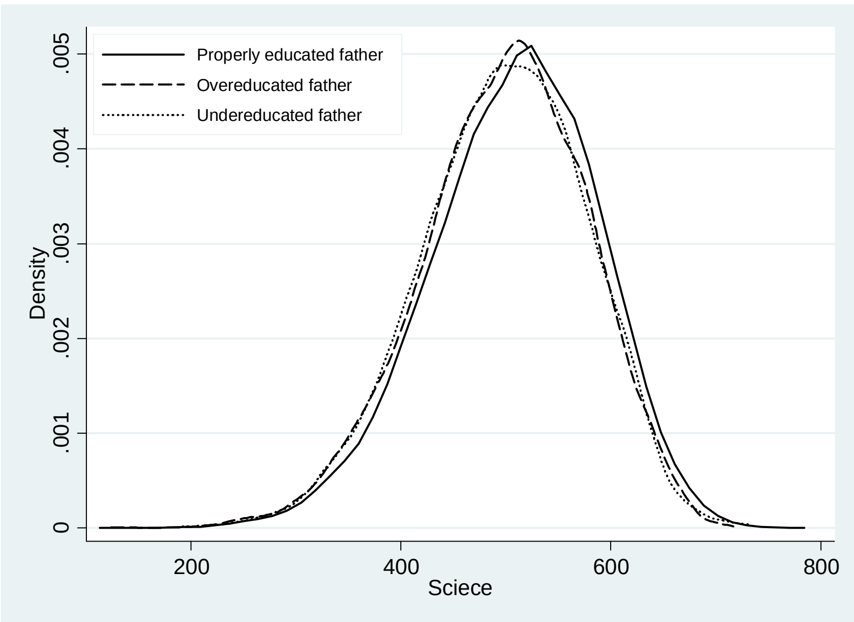
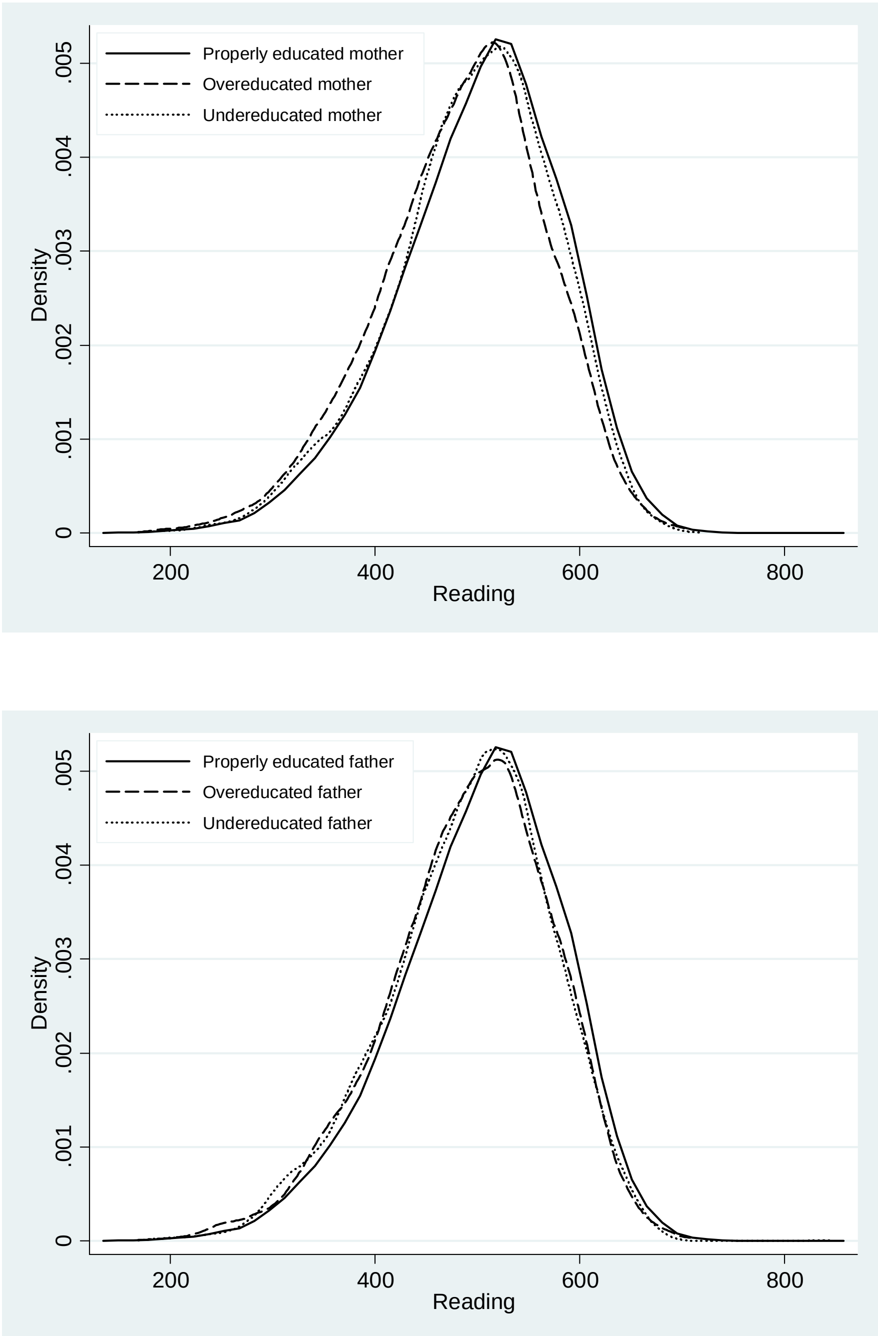


Figure 4.3. Students' educational outcomes in Reading according to parents' educational mismatch



4.5. Methodology and results

4.5.1. Methodology

To examine the relationship between parents' educational mismatch and students' educational outcomes, we follow the standard education production function (EPF) model defined by Levin (1974) and Hanushek (1979). The EPF models estimate outcomes – for instance, test scores – as a function of the cumulative influence of individual characteristics of the students, family and school inputs, and the peers of the student. Conceptually, the model is as equation (1):

$$RTest_i = \alpha + \beta \cdot IC_i + \delta \cdot FC_i + \gamma \cdot SC_i + u_i, \qquad (1)$$

where $RTest_i$ is achievement for student i, IC_i is a vector of students' individual characteristics, FC_i is a vector of family background influences, SC_i is vector of school inputs and u_i is an error term.

Following the model (1), we specify the following EPF model for each subject analysed by the PISA database:

$$\begin{aligned} RTest_i &= \alpha + \beta \cdot IC_i + \delta \cdot FC_i + \delta_1 \cdot MOver_i + \delta_2 \cdot FOver_i + \delta_3 \cdot MUnder_i + \\ &+ \delta_4 \cdot FUnder_i + \gamma \cdot SFE_i + u_i \end{aligned} \qquad (2)$$

where $RTest_i$ refers to the five plausible values of the test results for each subject (Mathematics, Science and Reading) of student *i* and IC_i and FC_i are vectors of control variables related to students and family background. The main difference between equation (2) and equation (1) is that equation (2) includes school fixed effects (SFE_i) instead of different variables related to school characteristics. As has been explained, we are not interested in the specific effect of school variables and, as schools are identified in the database, we include school fixed effects in the estimation. Thus, we control results via the heterogeneity among students derived from schools' differences.

In order to analyse the aims of the study, equation (2) also includes four variables relating to parents' educational mismatch. Specifically, it includes $MOver_i$, which takes the value of 1 if the student's mother is overeducated and 0 otherwise, and $FOver_i$, which takes the same values but in the case where the student's father is overeducated. On the other hand, $MUnder_i$ takes the value of 1 when the student's mother is undereducated and 0 otherwise, and the same values are for the variable $FUnder_i$ in case that father is undereducated. Finally, u_i is the error term.

Equation (2) is estimated taking into account specific characteristics of the PISA database[27]. On the one hand, we consider that the endogenous variable is based on five plausible values estimating the model for each plausible value by OLS. And, on the other hand, the complex sample design used in PISA requires applying a procedure based on replications to calculate the variance of the estimators. For this type of data, OECD (2009) recommends the method of balanced repeated replication (BRR) with a special modification of Fay (1989), which does not change the coefficients but improves the precision of the estimator of the variance. However, it is important to notice that quantile regressions are not allowed to be estimated taking into account the BRR method using standard procedures. So, we estimate the equation (2) with and without the BRR method in order to analyse the differences between both and to be able to compare the quantile regressions' results.

4.5.2. Results

Tables 4.2, 4.3 and 4.4 show the results of the OLS estimation of model (2) for the three different subjects (Mathematics, Sciences and Reading). Specifically, column (1) of Tables 4.2, 4.3 and 4.4 shows the estimation of the model without the variables related to the educational mismatch of parents, unlike column (2), which includes them. Both estimations consider the BBR method, which improves the precision of the estimator of the variance. Column (3) shows the results of the same estimation that is shown in column (2) without considering the BRR method. Finally, columns (4) to (8)

[27] The STATA's module, PV, is used: Stata module to perform estimation with plausible values. http://ideas.repec.org/c/boc/bocode/s456951.html

correspond to the model quantile estimates of column (3). All estimations include fixed effects for the 889 schools analysed.

Results of control variables largely support those obtained in previous studies[28]. As expected, students' gender has a statistically significant effect on students' educational achievement, and the sign is different depending on the subject analysed. Girls give a lower academic performance in Mathematics and Sciences than boys, but give a better academic performance in Reading. Furthermore, the quantile analysis shows that the higher the students' educational achievement, the greater the gap in educational achievement between girls and boys in Mathematics and Science. On the contrary, the higher outcomes girls have in Reading decrease along the distribution. The age of the students, between 15.3 and 16.3 years, has a positive impact as it increases in Sciences and Reading. As for students' nationality, first-generation immigrants suffer a penalty in their educational achievement in comparison to native students, although this penalty is lower for immigrants with better results. However, there is no significant effect on educational performance of students who are second-generation immigrants (i.e., those who were born in Spain) or whether the language of the PISA test is different from the primary language spoken in students' homes.

In regards to the set of variables related to students' family background, it is observed that the variables related to the number of books at home (included as an indicator of the cultural environment) have a positive effect on students' educational performance, which increases as the number of books in the home is greater. The type of family structure also has a significant effect on educational performance of students. In fact, having a single parent has a negative effect on student performance in Mathematics, but this effect is not statistically significant in either Sciences or Reading. However, having other family structures has a negative effect in all subjects compared to the nuclear family (consisting of both father and mother). On the other hand, it is generally found that mothers' educational level has a positive effect on students' educational performance, which becomes more important for students with a lower performance. However, this is not the case for fathers'

[28] Similar results are found in Woessmann (2003) using the TIMSS database and Fuchs and Woessmann (2007) using the PISA database.

educational level, since this variable is not statistically significant. The socioeconomic status of both parents is, in almost all cases, statistically significant, and the effect on educational performance of students is greater as the socioeconomic status increases.

As for the variables of interest in the study, we find that the educational performance of students whose parents are overeducated is lower than the educational performance of those students whose parents are properly educated. This negative effect is found in all three subjects tested, but the penalty is stronger in Mathematics than in Sciences or Reading. We also find that this effect is not equally distributed for all students. The quantile regressions show that the negative effect of parents' overeducation is much more intense for students with fewer educational outcomes for those who perform better on the three subjects tested. That is, students with higher educational performance are less affected by the overeducation of their parents. Results also point out that the relationship between undereducated parents and their children's educational performance has a different pattern depending on whether the mother or the father is undereducated. While there is no statistically significant relationship between undereducated fathers and the educational performance of their children, we find that the undereducation in mothers positively affects the educational performance of their children in Sciences (although no significant effect is appreciated in Mathematics and Reading). Quantile analysis shows no clear pattern of the effect of mothers' undereducation on students' performance, although it seems that the incidence is higher for students with lower educational outcomes in Sciences.

So, our hypothesis that parents' educational mismatch can alter their children's perception of the value of education seems to be supported by the results. In particular, the low educational achievement of students whose parents are overeducated may be indicating a "discouragement" effect derived from the experience of their parents. It is important to notice that overeducation has a negative effect on wages and job satisfaction of workers who are suffering from it, so children may have the perception that the opportunity cost of investing in education will exceed future profits obtained by increasing their levels of human capital. Furthermore, students who perform worse in school are more sensitive to being penalised by this intergenerational effect.

Regarding the relation between undereducated parents and their children's educational performance, obtained results are not so clear and conclusive. The positive relation between both variables may suggest that children could be aware of the troubles that their parents could suffer as a result of training deficiencies that would lead them to put more effort into their educational outcomes. However, this hypothesis is supported only in mothers and in one of the three subjects analysed.

To sum up, the results show that students whose parents are overeducated have lower educational performance in Mathematics, Sciences and Reading compared to students whose parents are properly educated, this effect being higher for low-performance students. However, it is important to highlight that the magnitude of the impact of parents' overeducation is lower than other personal and family characteristics.

4.5.3. Robustness checks

In order to check the robustness of our results, we carry out different extensions of our analysis.

First, we replicate all the estimations measuring parents' educational mismatch using the mean version of the statistical method instead of the mode version. Thus, we analyse whether the effects of parents' educational mismatch on students' outcomes also holds when the measurement method is changed. Although in a different context, Hartog (2000) points out that the incidence of overeducation is different depending on the measurement method used, although its effect on wages is quite robust and consistent: it is negative and statistically significant. We would like to test if something similar happens in the context of our research. As is shown in Table 1, the incidence of educational mismatch measured by the mean version is lower than the incidence obtained using the mode method. Tables A4.2–A4.4 in the Annex show the results of the estimations of the EPF for Mathematics, Sciences and Reading.

Table 4.2. OLS estimation of the EPF in Mathematics (continues)

				Quantile regression results				
Variables	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8) q90
Female	-22.16***	-22.06***	-22.06***	-11.80***	-14.47***	-17.55***	-20.88***	-22.78***
	(2.198)	(2.250)	(1.627)	(2.906)	(2.050)	(1.769)	(1.962)	(2.284)
Age	5.818	6.210	6.210**	12.19***	9.219*	7.848**	9.084**	11.20***
	(4.200)	(4.272)	(2.784)	(4.350)	(5.329)	(3.125)	(4.548)	(4.176)
Immigrant 1st generation	-18.82***	-17.25***	-17.25***	-29.03***	-25.94***	-20.18***	-15.42***	-7.184
	(5.818)	(5.843)	(3.800)	(7.395)	(4.954)	(4.413)	(4.966)	(7.358)
Immigrant 2nd generation	-4.029	-3.894	-3.894	-2.257	-6.947	-6.877	-5.508	9.328
	(12.28)	(12.38)	(7.865)	(17.98)	(11.73)	(11.36)	(14.62)	(15.46)
Language	0.834	1.231	1.231	-0.828	-0.221	-2.381	-2.936	-3.151
	(3.361)	(3.363)	(2.061)	(4.451)	(2.897)	(2.496)	(3.344)	(3.948)
Single parent family	-8.822**	-8.238**	-8.238***	-16.83***	-10.16***	-6.879**	-3.942	-3.826
	(3.912)	(3.932)	(2.682)	(5.866)	(3.931)	(3.025)	(3.241)	(3.961)
Mixed family	-41.11***	-39.91***	-39.91***	-29.40	-29.43**	-23.47**	-23.38**	-29.62**
	(12.20)	(12.25)	(8.527)	(21.36)	(13.89)	(11.25)	(10.46)	(13.80)
Education mother	0.152	0.967*	0.967***	0.833	0.707*	0.636	0.563	0.131
	(0.356)	(0.522)	(0.371)	(0.712)	(0.404)	(0.410)	(0.498)	(0.625)
Education father	-0.527	-0.205	-0.205	-0.0580	-0.0918	-0.287	-0.0708	-0.375
	(0.423)	(0.645)	(0.365)	(0.720)	(0.568)	(0.440)	(0.424)	(0.522)
ISEI mother	0.152*	0.00913	0.00913	0.107	0.101	0.130*	0.116	0.170*
	(0.0817)	(0.104)	(0.0636)	(0.109)	(0.0775)	(0.0709)	(0.0739)	(0.0913)
ISEI father	0.207**	0.0689	0.0689	0.147	0.101	0.114	0.0620	0.0910
	(0.0979)	(0.115)	(0.0713)	(0.122)	(0.104)	(0.0802)	(0.0831)	(0.127)
Overeducation mother		-9.426***	-9.426***	-12.19***	-10.08***	-9.026***	-7.326***	-5.800*
		(3.254)	(2.509)	(4.453)	(2.290)	(2.383)	(2.681)	(3.172)
Overeducation father		-10.70***	-10.70***	-9.844***	-10.92***	-7.895***	-6.139**	-4.310
		(3.777)	(2.051)	(3.559)	(3.485)	(3.010)	(2.769)	(3.326)

Table 4.2. (continuation): OLS estimation of the EPF in Mathematics

Variables	(1)	(2)	(3)	Quantile regression results (4) q10	(5) q25	(6) q50	(7) q75	(8) q90
Undereducation mother		2.297	2.297	3.793	2.999	3.175	4.066	2.786
		(3.644)	(2.658)	(4.290)	(2.870)	(2.218)	(2.747)	(3.769)
Undereducation father		-4.561	-4.561	-3.757	-2.542	-3.286	-1.266	-1.847
		(3.849)	(2.847)	(5.115)	(3.951)	(3.171)	(2.946)	(3.685)
From 11 to 25 books	20.05***	19.62***	19.62***	19.46**	21.16***	21.23***	16.98**	16.78**
	(7.206)	(7.149)	(3.366)	(7.891)	(6.347)	(5.808)	(6.735)	(7.184)
From 26 to 100 books	46.41***	46.23***	46.23***	49.47***	51.08***	49.71***	43.66***	41.13***
	(7.617)	(7.509)	(2.989)	(8.124)	(5.274)	(5.248)	(6.371)	(7.365)
From 101 to 200 books	67.86***	67.38***	67.38***	68.90***	69.95***	67.13***	60.30***	55.73***
	(7.477)	(7.319)	(3.187)	(9.837)	(5.742)	(5.387)	(6.784)	(8.568)
From 201 to 500 books	79.68***	79.02***	79.02***	80.54***	80.47***	77.10***	72.60***	69.28***
	(7.651)	(7.556)	(3.567)	(8.488)	(5.881)	(5.663)	(7.445)	(9.093)
More than 500 books	81.69***	80.63***	80.63***	77.98***	81.24***	79.85***	75.24***	71.04***
	(8.119)	(7.913)	(3.796)	(7.925)	(6.264)	(5.891)	(7.452)	(8.772)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-144.8**	-147.0**	-147.0***	-344.0***	-248.4***	-176.1***	-144.6**	-132.6*
	(65.90)	(68.23)	(45.99)	(72.93)	(84.57)	(52.22)	(72.17)	(67.80)
Average R-squared	0.1595	0.1644	0.1644					
Observations	13098	13098	13098	13098	13098	13098	13098	13098

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

Table 4.3. OLS estimation of the EPF in Sciences (continues)

				Quantile regression results				
Variables	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8) q90
Female	-13.99***	-13.94***	-13.94***	-5.272**	-8.099***	-12.59***	-16.28***	-18.37***
	(2.124)	(2.149)	(1.453)	(2.520)	(2.104)	(1.836)	(1.767)	(3.277)
Age	15.14***	15.35***	15.35***	14.01***	12.22***	11.05***	9.095***	9.541**
	(3.563)	(3.616)	(2.357)	(3.857)	(3.217)	(3.025)	(2.957)	(4.697)
Immigrant 1st generation	-16.56***	-15.56***	-15.56***	-21.22***	-21.05***	-14.86***	-12.26**	-4.639
	(5.313)	(5.364)	(3.827)	(5.740)	(4.487)	(3.609)	(4.804)	(6.631)
Immigrant 2nd generation	-0.494	-0.0856	-0.0856	-1.364	-6.713	-6.475	-5.990	-8.896
	(10.10)	(10.18)	(6.731)	(16.65)	(9.014)	(9.795)	(11.65)	(10.70)
Language	0.366	0.722	0.722	-5.103	-2.243	-1.903	-2.727	-3.670
	(3.032)	(3.025)	(2.059)	(3.897)	(2.764)	(3.260)	(3.664)	(4.881)
Single parent family	-3.709	-3.276	-3.276	-9.343*	-4.533	-1.290	-1.320	-1.536
	(3.466)	(3.454)	(2.213)	(5.133)	(3.711)	(3.430)	(3.080)	(3.676)
Mixed family	-24.94*	-24.11*	-24.11***	-32.87*	-16.52	-13.90	-0.635	5.102
	(13.14)	(12.87)	(9.227)	(19.83)	(14.57)	(12.48)	(14.30)	(14.56)
Education mother	0.542	1.674***	1.674***	1.502***	1.561***	1.171***	0.971*	0.687
	(0.373)	(0.528)	(0.338)	(0.578)	(0.407)	(0.382)	(0.501)	(0.569)
Education father	-0.598*	-0.538	-0.538	-0.289	-0.408	-0.507	-0.514	-0.329
	(0.362)	(0.619)	(0.370)	(0.649)	(0.559)	(0.430)	(0.372)	(0.532)
ISEI mother	0.241***	0.0755	0.0755	0.0395	0.0323	0.0818	0.0960	0.109
	(0.0713)	(0.101)	(0.0657)	(0.0971)	(0.0818)	(0.0743)	(0.0943)	(0.106)
ISEI father	0.0492	-0.0219	-0.0219	0.129	0.112	0.125	0.124	0.0785
	(0.0850)	(0.113)	(0.0721)	(0.143)	(0.120)	(0.0908)	(0.0781)	(0.0994)
Overeducation mother		-7.737**	-7.737***	-7.617*	-8.702***	-8.868***	-7.026**	-4.107
		(3.426)	(2.447)	(3.927)	(2.677)	(2.090)	(2.889)	(3.249)
Overeducation father		-6.228*	-6.228***	-5.528	-5.725*	-5.473**	-3.954	-2.654
		(3.417)	(2.160)	(4.038)	(3.181)	(2.400)	(2.418)	(3.555)

Table 4.3. (continuation): OLS estimation of the EPF in Sciences

				Quantile regression results				
Variables	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8) q90
Undereducation mother		6.594*	6.594***	4.648	4.049*	4.374*	4.052	3.935
		(3.498)	(2.256)	(3.745)	(2.377)	(2.468)	(3.274)	(3.446)
Undereducation father		-3.715	-3.715	-1.607	-1.227	-3.549	-2.481	-1.135
		(3.487)	(2.428)	(3.545)	(3.354)	(2.578)	(2.902)	(3.772)
From 11 to 25 books	16.05***	15.40**	15.40***	20.83***	17.42***	14.99***	13.83***	14.58*
	(5.945)	(5.992)	(3.301)	(7.584)	(6.650)	(4.599)	(4.687)	(7.651)
From 26 to 100 books	40.92***	40.31***	40.31***	46.09***	42.73***	39.18***	39.05***	38.30***
	(6.233)	(6.245)	(3.283)	(6.966)	(5.571)	(3.840)	(4.560)	(6.629)
From 101 to 200 books	58.45***	57.76***	57.76***	62.51***	58.50***	55.78***	55.15***	53.67***
	(6.230)	(6.175)	(3.788)	(6.903)	(5.522)	(4.069)	(4.738)	(7.840)
From 201 to 500 books	71.39***	70.39***	70.39***	73.93***	68.77***	64.10***	64.82***	65.89***
	(6.539)	(6.449)	(4.101)	(8.024)	(5.440)	(4.659)	(5.267)	(8.663)
More than 500 books	78.70***	77.58***	77.58***	73.96***	74.16***	71.10***	72.33***	71.16***
	(6.742)	(6.655)	(3.891)	(7.957)	(5.654)	(4.283)	(5.819)	(7.237)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-291.1***	-295.9***	-295.9***	-372.5***	-295.6***	-225.9***	-146.6***	-111.9
	(56.82)	(58.08)	(38.99)	(60.74)	(50.14)	(47.90)	(48.66)	(75.59)
Average R-squared	0.1449	0.1487	0.1487					
Observations	13098	13098	13098	13098	13098	13098	13098	13098

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

Table 4.4. OLS estimation of the EPF in Reading (continues)

				Quantile regression results				
Variables	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8) q90
Female	20.64***	20.73***	20.73***	32.45***	28.13***	22.80***	19.00***	16.46***
	(1.841)	(1.865)	(1.267)	(2.273)	(1.999)	(1.662)	(1.622)	(2.391)
Age	11.89***	12.14***	12.14***	13.87***	10.93***	11.59***	11.23***	11.14***
	(3.886)	(3.942)	(2.671)	(5.224)	(3.044)	(2.704)	(2.737)	(4.156)
Immigrant 1st generation	-19.51***	-18.50***	-18.50***	-22.69***	-18.45***	-15.37***	-10.06**	-4.471
	(4.823)	(4.767)	(3.151)	(6.191)	(5.029)	(4.206)	(4.292)	(5.581)
Immigrant 2nd generation	1.536	1.918	1.918	3.409	0.00295	-3.038	-0.376	2.183
	(10.13)	(10.15)	(6.605)	(17.28)	(10.42)	(9.323)	(11.71)	(14.34)
Language	-2.214	-1.922	-1.922	-8.653*	-6.971**	-3.926	-3.761	-2.612
	(2.774)	(2.787)	(1.851)	(5.034)	(2.920)	(2.769)	(2.973)	(3.216)
Single parent family	-4.782	-4.398	-4.398**	-8.960**	-4.314	-1.932	-1.780	-0.157
	(3.285)	(3.291)	(1.998)	(4.239)	(3.435)	(2.980)	(2.755)	(3.671)
Mixed family	-26.63*	-25.94*	-25.94***	-32.83*	-27.36**	-16.84	-12.81	-11.70
	(15.15)	(15.00)	(9.023)	(18.85)	(13.75)	(11.50)	(12.88)	(15.09)
Education mother	0.278	1.105**	1.105***	1.453***	0.906**	0.735*	0.576	0.329
	(0.352)	(0.499)	(0.298)	(0.544)	(0.450)	(0.401)	(0.379)	(0.508)
Education father	-0.679*	-0.437	-0.437	-0.116	-0.556	-0.425	-0.220	-0.455
	(0.366)	(0.571)	(0.328)	(0.592)	(0.430)	(0.335)	(0.397)	(0.541)
ISEI mother	0.194**	0.0716	0.0716	-0.0191	0.0570	0.0637	0.0508	0.0778
	(0.0783)	(0.0988)	(0.0566)	(0.0907)	(0.0732)	(0.0539)	(0.0683)	(0.0864)
ISEI father	0.179**	0.0770	0.0770	0.175	0.217**	0.152**	0.126	0.172*
	(0.0832)	(0.107)	(0.0655)	(0.131)	(0.0977)	(0.0774)	(0.0828)	(0.0994)
Overeducation mother		-5.928**	-5.928***	-10.64**	-7.391**	-7.969***	-6.520***	-3.807
		(2.674)	(2.021)	(4.161)	(2.966)	(2.340)	(2.453)	(3.540)
Overeducation father		-7.993**	-7.993***	-6.667*	-6.325**	-5.039**	-4.264	-0.477
		(3.280)	(2.043)	(3.755)	(2.951)	(2.231)	(2.765)	(3.262)

Table 4.4. (continuation): OLS estimation of the EPF in Reading

				Quantile regression results				
Variables	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8) q90
Undereducation mother		4.615	4.615**	3.950	3.604	3.315	4.395*	4.231
		(3.068)	(1.941)	(4.799)	(2.491)	(2.487)	(2.250)	(3.101)
Undereducation father		-3.245	-3.245*	-0.635	-2.876	-3.047	-2.466	-0.654
		(2.901)	(1.882)	(3.767)	(3.043)	(2.521)	(2.694)	(3.276)
From 11 to 25 books	16.63***	16.18***	16.18***	17.69***	19.57***	20.26***	14.05**	16.40***
	(6.177)	(6.173)	(3.889)	(6.640)	(5.035)	(6.887)	(6.069)	(6.333)
From 26 to 100 books	42.43***	42.02***	42.02***	45.61***	44.33***	42.14***	36.77***	38.01***
	(6.161)	(6.114)	(3.163)	(6.190)	(4.255)	(5.645)	(5.670)	(5.379)
From 101 to 200 books	58.48***	57.94***	57.94***	59.22***	57.37***	57.13***	51.34***	52.21***
	(5.818)	(5.720)	(3.483)	(6.755)	(4.912)	(6.172)	(6.211)	(5.770)
From 201 to 500 books	67.71***	66.96***	66.96***	71.82***	68.86***	67.22***	62.11***	63.92***
	(6.294)	(6.194)	(3.781)	(7.353)	(4.997)	(7.059)	(6.500)	(6.045)
More than 500 books	71.36***	70.39***	70.39***	68.41***	69.14***	70.95***	65.37***	67.56***
	(6.518)	(6.417)	(4.247)	(8.856)	(5.353)	(7.109)	(6.954)	(5.935)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-256.4***	-260.4***	-260.4***	-391.7***	-291.8***	-249.3***	-191.4***	-149.0**
	(62.02)	(62.98)	(42.81)	(81.95)	(49.77)	(43.08)	(44.28)	(65.99)
Average R-squared	0.1471	0.1502	0.1502					
Observations	13098	13098	13098	13098	13098	13098	13098	13098

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

We can see that the significance level of the variables of interest reduces or even disappears in most of the cases (see column (2) which shows the more efficient results of the EPF estimation, i.e., considering the BRR option). In particular, mother's overeducation does not have a statistically significant effect on students' outcomes, but, the effect exists in the case of overeducation in fathers[29]. So, the results obtained previously regarding the effect of parents' educational mismatch on students' outcomes should be carefully interpreted.

Second, we are aware that our results could be affected by a possible selection bias problem regarding parents' participation in the labour market (Heckman, 1979). In other words, the sample of students whose parents are working (and so, the ones for whom we could calculate job-education mismatch) may not be a suitable representation of all students. Unfortunately, the nature of the PISA database does not allow us to control for a selectivity problem, because there is not enough parent information. Then, in order to give some information about whether students whose parents are working perform in a different way than students whose parents are not working, we estimate an EPF considering the entire initial sample and including two dummy variables related to the labour status of parents. The first one is "Mother working", which takes the value 1 if the mother is working and 0 otherwise, and the second is "Father working", which takes the value 1 if father is working. The results (see Table A4.5 in the Annex) do not show significant differences in students' performance due to the labour status of their parents. So, this means that the effect of educational mismatch we found in our previous analysis does not seem to be capturing the effect of employment participation, although we recognise that more detailed analysis would be required in order to distinguish between the two effects, something we cannot do with the database considered here.

[29] In Mathematics, the effect is statistically significant at 17%, in Sciences, at a 5% significance level, and in Reading, the p-value is 12%.

4.6. Conclusions

This chapter analyses the relationship between parents' educational mismatch and the educational outcomes of their children in Mathematics, Sciences and Reading using the PISA 2009 microdata for Spain. The results from the estimation of an education production function show that students whose parents are overeducated have a lower educational performance in all three subjects tested in comparison to students whose parents are properly educated. Moreover, this penalty is stronger for students with a lower educational outcome. On the contrary, parents' undereducation has an impact only on educational performance of students in Sciences and only for undereducated mothers, being a positive effect. However, the intensity of the impact of parents' overeducation on educational performance of their children is not relatively high. Variables such as gender, age, nationality, type of family structure, or cultural environment have a greater impact on students' performance.

It is worth highlighting that this analysis shows only the existence of a relationship between educational mismatch of parents and educational performance of their children, but we cannot talk about causality between the two variables considered without talking about the direction of it. In order to distinguish between causality and association, it would be necessary to have a richer database in regards to information on both families and students. Such information should, first, allow for identification of the mechanisms through which the educational mismatch of parents ends up having effects on their children's educational performance and to what extent this effect is not capturing the importance of other variables that we could not control adequately in this study (omitted variables). In this sense, the hypothesis that has been formulated in this study is based only on the perception of the students about the importance of human capital in the parents' job performance, while other channels could exist linked – for example, to less dedication to the children resulting from the need to work longer hours to get the same level of income, etc.

In any case, the results presented in the paper provide additional arguments to take policy measures in order to reduce overeducation. In particular, besides the findings of the previous analysis about the negative consequences of overeducation on wages (see Chapter 2 and Chapter 3 of the PhD dissertation; Groot and Maassen van der Brink, 2000), productivity (Tsang et.al., 1991) and job satisfaction (Fleming and Kler, 2008), our results point out that overeducation may also have negative effects on educational outcomes of children.

References

Ammermueller, A. (2007) "Poor background or low returns? Why immigrant students in Germany perform so poorly in the programme for international student assessment", *Education Economics*, 15(2), 215-230.

Barnett, R. R.; Glass, J. C.; Snowdon, R. I. and Stringer, K. S. (2002) "Size, performance and effectiveness: cost-constrained measures of best-practice performance and secondary-school size", *Education Economics*, 10 (3), 291-311.

Björklund, A. and Salvanes, K. G. (2010) "Education and Family Background: Mechanisms and Policies", in E.A. Hanushek, S. Machin, and L. Woessmann (eds.), *Handbook in Economics of Education*, Vol. 3, North Holland: 2010, 201-247.

Calero, J. and Escardíbul, J.O. (2007) "Evaluación de servicios educativos: el rendimiento en los centros públicos y privados medido en PISA-2003", *Hacienda Pública Española / Revista de Economía Pública*, 183 (4), 33-66.

Calero, J. and Waisgrais, S. (2008) "¿Qué determina el rendimiento de los alumnos inmigrantes? Una primera aproximación a partir de PISA-06", *Investigaciones de Economía de la Educación*, 3, 499-508.

Coleman, J. S. and Hoffer, T. (1987) *Public and private high schools. The impact of communities.* New York: Basic Books.

Coleman, J. S.; Campbell, J. S.; Campbell, E. Q.; Hobson, C. J.; McPartland, J.; Mood, A. M.; Weinfeld, F. D. and York, R. L. (1966) *Equality of Educational Opportunity*. Washington, D.C.: US Department of Health, Education, and Welfare, US Government Printing Office.

Dahl, G.B. and Lochner, L. (2012) "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit", *American Economic Review*, 102(5), 1927-1956.

Entorf, H. and Lauk, M. (2008) "Peer effects, social multipliers and migrants at school: An international comparison", *Journal of Ethnic and Migration Studies*, 34(4), 633-654.

Fay, R. E. (1989) "Theoretical application of weighting for variance calculation", in *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 212-217.

Feinstein, L. and Symons, J. (1999) "Attainment in secondary education", *Oxford Economic Papers*, 51, 300-321.

Fleming, C. M. and Kler, P. (2008) "I'm too clever for this job: a bivariate probit analysis on overeducation and job satisfaction in Australia". *Applied Economics*, 40 (9), 1123-1138.

Fuchs, T. and Woessmann, L. (2007) "What accounts for international differences in students' performance? A re-examination using PISA data", *Empirical Economics*, 32, 433-464.

Ganzeboom, H.; De Graaf, P.; Treiman, J. and De Leeuw, J. (1992) "A standard international socio-economic index of occupational status", *Social Science Research*, 21(1), 1-56.

Groot, W. and Maassen van den Brink, H. (2000) "Overeducation in the Labor Market: a Meta-Analysis", *Economics of Education Review*, 19(2), 149-158.

Häkkinen, I,; Kirjavainen, T. and Uusitalo, R. (2003) "School resources and student achievement revisited: new evidence from panel data", *Economics of Education Review*, 22, 329-335.

Hampden-Thomson, G. (2013) "Family policy, family structure, and children's educational achievement", *Social Science Research*, 42, 804-817.

Hanushek, E. A. (1979) "Conceptual and empirical issues in the estimation of educational production functions", *Journal of Human Resources*, 14, 351-388.

Hanushek, E. A. (1986) "The economics of schooling", *Journal of Economic Literature*, 24, 1141-1177.

Hanushek, E. A. (2003) "The failure of input based schooling policies", *The Economic Journal*, 113, 64-98.

Hanushek, E. A. and Luque, J. A. (2003) "Efficiency and equity in schools around the world", *Economics of Education Review*, 22, 481-502.

Hartog, J. (2000) "Over-education and earnings: Where are we, where should we go?", *Economics of Education Review*, 19, 131-147.

Haveman, R. and Wolfe, B. (1995) "The determinants of children's attainment: A review of methods and findings", *Journal of Economics Literature*, 33(4), 1829-1878.

Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-161.

Kiker, B. F.; Santos, M. C. and Mendes de Oliveira, M. (1997) "Overeducation and Undereducation: Evidence for Portugal", *Economics of Education Review*, 16(2), 111-125.

Krueger, A. B. (2003) "Economics considerations and class size", *Economic Journal*, 113, 34-63.

Leuven, E. and Oosterbeek, H. (2011) "Overeducation and Mismatch in the Labor Market", in: E. Hanushek, S. Machin, L. Woessman (Eds.), *Handbook of the Economics of Education*, 4, 283-326.

Levin, H.M. (1974) "Measuring efficiency in educational production", *Public Finance Quartely*, 2, 3-24.

Marks, G.N. (2008) "Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries", *Oxford Review of Education*, 34(1), 89-109.

Martins, L. and Veiga, V. (2010) "Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities?", *Economics of Education Review*, 29 (6), 1016-1033.

Meunier, M. (2011) "Immigration and student achievement: Evidence from Switzerland", *Economics of Education Review*, 30 (1), 16-38.

Mislevy, R. J. and Sheehan, K. M. (1987) "Marginal estimation procedures", in Beaton, A.E., (Ed.), *The NAEP 1983-84 technical report*, National

Assessment of Educational Progress, Educational Testing Service, Princeton, 293-360.

Mislevy, R. J. and Sheehan, K. M. (1989) "Information matrices in latent-variable models", *Journal of Educational Statistics*, 14, 335-350.

Noell, J. (1982) "Public and Catholic schools: A re-analysis of public and private schools", *Sociology of Education*, 55, 123-132.

OECD (2009) *Technical report 2006*, Paris.

OECD (2010) *PISA 2009 Results: Learning Trends: Changes in Student Performance Since 2000 (Volume V)*, Paris.

Opdenakker, M. C. and Van Damme, J. (2006) "Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools", *School Effectiveness and School Improvement*, 17(1), 87-117.

Plug, E. and Vijverberg, W. (2003) "Schooling, Family Background, and Adoption: Is It Nature or Is It Nurture?", *Journal of Political Economy,* 111(3), 611-641.

Perelman, S. and Santin, D. (2011) "Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results", *Education Economics*, 19(1), 29-49.

Quintini, G. (2011) "Over-Qualified or Under-Skilled: A Review of Existing Literature", *OECD Social, Employment and Migration Working Papers*, No. 121, OECD Publishing.

Rivkin, S. G.; Hanushek, E. A. and Kain, J. F. (2005) "Teachers, Schools and Academic Achievement". *Econometrica*, 73(2), 417-458.

Rubin, D. B. (1987) *Multiple imputation for nonresponse in surveys*, New York, Wiley.

Sander, W. (1996) "Catholic grade schools and academic achievement", *The Journal of Human Resources*, 31 (3), 540-548.

Smith, J. and Naylor, R.A. (2005) "Schooling effects on subsequent university performance: evidence for the UK university population", *Economics of Education Review*, 24, 549-562.

Stevans, L. K. and Sessions, D. N. (2000) "Private/public school choice and student performance revisited", *Education Economics*, 8(2), 169-184.

Tsang, M. C.; Rumberger, R. and Levin, H.M. (1991) "The impact of surplus schooling on worker productivity", *Industrial Relations*, 30(2), 209-228.

Verdugo, R. and Verdugo, N. (1989) "The impact of surplus schooling on earnings", *Journal of Human Resources*, 24(4), 629-643.

Woessmann, L. (2003) "Schooling resources, educational institutions and student performance: the international evidence", *Oxford Bulletin of Economics and Statistics*, 65(2), 117-170.

Annex chapter 4

Table A4.1. Descriptive statistics

	Variable	Mean	Std. Dev.	Min	Max
Scores:	Mathematics	500.8207	82.37742	126.662	789.54
	Sciences	502.9138	79.85539	124.058	773.816
	Reading	496.8703	79.78956	144.992	847.098
Individual characteristics:	Female	0.4923198	0.4999601	0	1
	Age	15.85666	0.2866355	15.33	16.33
	Natives	0.9170093	0.2758787	0	1
	Immigrant 1st generation	0.073422	0.2608379	0	1
	Immigrant 2nd generation	0.0095687	0.0973544	0	1
	Language at home is language of test	0.8156116	0.3878155	0	1
	Language at home is different as language of test	0.1843884	0.3878155	0	1
Family characteristics:	Nuclear family	0.8796026	0.325438	0	1
	Single parent family	0.1137906	0.3175689	0	1
	Mixed	0.0066067	0.0810159	0	1
	Education mother	11.93948	3.905337	3	16.5
	Education father	11.44388	4.110173	3	16.5
	ISEI mother	42.57494	18.47744	16	90
	ISEI father	44.64007	16.82012	16	90
	Books at home: from 0 to 10 books	0.0551942	0.2283677	0	1
	from 11 to 25	0.1127449	0.3162928	0	1
	from 26 to 100	0.3015851	0.4589636	0	1
	from 101 to 200	0.2258476	0.4181552	0	1
	from 201 to 500	0.1788927	0.3832771	0	1
	more than 500	0.1257355	0.3315636	0	1
Regions:	Andalusia	0.1713198	0.3768026	0	1
	Aragón	0.028884	0.167487	0	1
	Asturias	0.0170594	0.1294978	0	1
	Balearic Islands	0.0262761	0.1599614	0	1
	Canary Islands	0.0381441	0.1915513	0	1
	Cantabria	0.0110565	0.1045709	0	1
	Castile and León	0.0485259	0.2148829	0	1
	Catalonia	0.1688073	0.3745959	0	1
	Galicia	0.0562862	0.2304823	0	1
	Rioja	0.0067338	0.0817858	0	1
	Madrid	0.150378	0.3574552	0	1
	Murcia	0.032203	0.1765457	0	1
	Navarra	0.0151029	0.1219669	0	1
	Basque Country	0.0470182	0.2116858	0	1
	Ceuta and Melilla	0.0029532	0.0542653	0	1
	Not awarded region	0.1792516	0.3835775	0	1

Note: 13098 observations

Table A4.2. OLS estimation of the EPF in Mathematics using mean method (continues)

				Quantile regression results				
	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8)q90
Female	-22.16***	-22.22***	-22.22***	-11.64***	-14.18***	-17.29***	-21.04***	-22.57***
	(2.198)	(2.178)	(1.613)	(2.776)	(2.139)	(1.603)	(1.935)	(2.142)
Age	5.818	5.951	5.951**	11.63***	9.317**	8.276**	9.592**	11.16**
	(4.200)	(4.226)	(2.811)	(3.926)	(4.667)	(3.758)	(4.137)	(4.654)
Immigrant 1st generation	-18.82***	-18.81***	-18.81***	-28.69***	-25.99***	-21.02***	-15.59***	-7.355
	(5.818)	(5.758)	(3.646)	(6.668)	(5.145)	(5.223)	(5.291)	(6.350)
Immigrant 2nd generation	-4.029	-2.936	-2.936	-2.392	-3.866	-7.688	-4.071	10.46
	(12.28)	(12.34)	(7.814)	(16.54)	(11.52)	(10.57)	(12.86)	(14.50)
Language	0.834	1.148	1.148	-2.168	-0.563	-1.995	-3.293	-3.394
	(3.361)	(3.346)	(2.057)	(4.232)	(3.888)	(2.495)	(3.252)	(3.845)
Single parent family	-8.822**	-8.716**	-8.716***	-16.88***	-10.57***	-6.820**	-4.080	-3.802
	(3.912)	(3.931)	(2.658)	(4.761)	(3.537)	(2.792)	(3.049)	(3.782)
Mixed family	-41.11***	-40.36***	-40.36***	-31.27	-28.72*	-22.63**	-22.41*	-27.25**
	(12.20)	(12.23)	(8.465)	(23.62)	(15.28)	(11.39)	(11.47)	(13.12)
Education mother	0.152	0.439	0.439	0.136	0.633	0.549	0.456	0.142
	(0.356)	(0.556)	(0.385)	(0.825)	(0.647)	(0.492)	(0.623)	(0.640)
Education father	-0.527	-0.506	-0.506	-0.218	-0.168	-0.362	-0.459	-0.778
	(0.423)	(0.607)	(0.367)	(0.732)	(0.516)	(0.458)	(0.496)	(0.590)
ISEI mother	0.152*	0.134	0.134**	0.238**	0.186*	0.215***	0.186**	0.211**
	(0.0817)	(0.0898)	(0.0573)	(0.113)	(0.100)	(0.0703)	(0.0843)	(0.100)
ISEI father	0.207**	0.176	0.176**	0.273**	0.200*	0.178**	0.150*	0.175*
	(0.0979)	(0.109)	(0.0719)	(0.114)	(0.105)	(0.0737)	(0.0786)	(0.102)
Overeducation mother		-0.776	-0.776	-6.520	-8.111**	-7.245**	-4.896	-4.108
		(3.676)	(2.300)	(5.184)	(3.328)	(3.585)	(3.786)	(4.534)
Overeducation father		-6.141	-6.141**	-5.780	-7.487*	-4.781*	-1.753	-0.572
		(4.428)	(2.713)	(5.147)	(4.156)	(2.628)	(3.460)	(5.137)
Undereducation mother		3.386	3.386	0.334	4.894	4.976	6.010	5.698
		(4.671)	(3.083)	(6.110)	(3.826)	(3.289)	(4.450)	(4.364)
Undereducation father		-4.732	-4.732	-2.652	-1.077	-1.130	-2.794	-4.906
		(4.825)	(3.177)	(6.790)	(4.186)	(4.204)	(3.820)	(4.450)

Table A4.2. (continuation): OLS estimation of the EPF in Mathematics using mean method

				Quantile regression results				
	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8)q90
From 11 to 25 books	20.05***	19.96***	19.96***	20.35***	21.60***	21.34***	17.16**	16.29**
	(7.206)	(7.129)	(3.344)	(7.749)	(7.442)	(5.856)	(7.147)	(8.308)
From 26 to 100 books	46.41***	46.27***	46.27***	49.91***	51.10***	49.83***	44.02***	41.08***
	(7.617)	(7.520)	(3.016)	(8.334)	(6.927)	(5.604)	(5.951)	(7.424)
From 101 to 200 books	67.86***	67.76***	67.76***	69.82***	70.33***	67.33***	60.93***	55.79***
	(7.477)	(7.373)	(3.202)	(8.680)	(7.204)	(5.040)	(6.910)	(8.849)
From 201 to 500 books	79.68***	79.54***	79.54***	81.67***	80.54***	77.30***	72.88***	69.10***
	(7.651)	(7.573)	(3.581)	(8.787)	(7.177)	(5.908)	(7.680)	(8.458)
More than 500 books	81.69***	81.70***	81.70***	80.41***	81.20***	80.23***	76.02***	71.62***
	(8.119)	(8.007)	(3.822)	(8.630)	(7.811)	(5.970)	(7.856)	(7.520)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-144.8**	-147.1**	-147.1***	-341.4***	-259.7***	-190.2***	-155.2**	-134.4*
	(65.90)	(66.76)	(46.43)	(65.37)	(77.18)	(61.04)	(66.65)	(74.31)
Average R-squared	0.1595	0.1608	0.1608					
Observations	13098	13098	13098	13098	13098	13098	13098	13098

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

Table A4.3. OLS estimation of the EPF in Sciences using mean method (continues)

				Quantile regression results				
	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8)q90
Female	-13.99***	-14.02***	-14.02***	-5.515**	-8.318***	-12.62***	-16.27***	-18.44***
	(2.124)	(2.115)	(1.444)	(2.344)	(1.886)	(1.770)	(1.628)	(3.104)
Age	15.14***	15.25***	15.25***	14.00***	11.71***	10.17***	9.081**	8.854*
	(3.563)	(3.563)	(2.376)	(3.862)	(2.963)	(2.746)	(3.639)	(4.625)
Immigrant 1st generation	-16.56***	-16.52***	-16.52***	-20.76***	-20.43***	-16.63***	-12.93***	-5.543
	(5.313)	(5.260)	(3.809)	(5.266)	(4.718)	(4.477)	(4.557)	(7.248)
Immigrant 2nd generation	-0.494	0.445	0.445	-0.0103	-5.386	-7.661	-5.790	-10.55
	(10.10)	(10.01)	(6.749)	(17.12)	(8.667)	(9.803)	(11.73)	(9.911)
Language	0.366	0.748	0.748	-4.000	-2.894	-2.125	-2.652	-4.441
	(3.032)	(2.979)	(2.078)	(4.008)	(2.902)	(2.832)	(3.246)	(3.766)
Single parent family	-3.709	-3.620	-3.620	-10.47**	-5.353*	-0.867	-1.132	-1.571
	(3.466)	(3.478)	(2.216)	(5.131)	(2.896)	(3.362)	(3.377)	(3.774)
Mixed family	-24.94*	-24.11*	-24.11***	-34.39*	-15.11	-14.04	-1.868	4.517
	(13.14)	(13.04)	(9.219)	(19.74)	(13.89)	(12.59)	(12.63)	(13.21)
Education mother	0.542	0.573	0.573	0.817	1.266**	0.938**	0.571	0.344
	(0.373)	(0.547)	(0.349)	(0.672)	(0.540)	(0.450)	(0.500)	(0.525)
Education father	-0.598*	-0.330	-0.330	-0.172	-0.387	-0.278	-0.260	-0.236
	(0.362)	(0.595)	(0.416)	(0.803)	(0.560)	(0.397)	(0.479)	(0.555)
ISEI mother	0.241***	0.241***	0.241***	0.158*	0.122	0.169**	0.183*	0.188**
	(0.0713)	(0.0805)	(0.0550)	(0.0879)	(0.0840)	(0.0697)	(0.0956)	(0.0870)
ISEI father	0.0492	-0.00946	-0.00946	0.148	0.162	0.131*	0.126	0.0844
	(0.0850)	(0.112)	(0.0731)	(0.128)	(0.104)	(0.0673)	(0.0911)	(0.0963)
Overeducation mother		-0.318	-0.318	-5.289	-7.575**	-6.729**	-3.917	-2.023
		(3.331)	(2.268)	(4.229)	(3.076)	(2.773)	(3.285)	(3.362)
Overeducation father		-7.369*	-7.369***	-7.560*	-6.015*	-6.878***	-5.185	-3.261
		(4.028)	(2.519)	(3.909)	(3.415)	(2.293)	(3.410)	(2.969)
Undereducation mother		0.275	0.275	-1.250	2.795	4.759	3.516	2.006
		(3.771)	(2.467)	(6.431)	(3.757)	(2.907)	(3.184)	(3.711)
Undereducation father		-2.422	-2.422	-0.700	-1.172	-2.068	-1.233	-0.902
		(3.689)	(2.732)	(5.884)	(4.338)	(3.419)	(3.731)	(4.150)

Table A4.3. (continuation): OLS estimation of the EPF in Sciences using mean method

				Quantile regression results				
	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8)q90
From 11 to 25 books	16.05***	16.08***	16.08***	21.15**	17.26**	15.37***	13.26***	13.74*
	(5.945)	(5.867)	(3.309)	(8.704)	(7.286)	(4.925)	(4.898)	(7.168)
From 26 to 100 books	40.92***	40.82***	40.82***	45.79***	42.71***	39.25***	38.49***	37.86***
	(6.233)	(6.167)	(3.317)	(7.212)	(6.658)	(3.937)	(4.520)	(6.214)
From 101 to 200 books	58.45***	58.33***	58.33***	63.21***	58.37***	56.19***	54.88***	53.23***
	(6.230)	(6.144)	(3.811)	(7.880)	(6.690)	(4.247)	(4.435)	(6.914)
From 201 to 500 books	71.39***	71.27***	71.27***	74.32***	69.26***	64.45***	64.90***	64.98***
	(6.539)	(6.504)	(4.156)	(7.941)	(7.093)	(4.533)	(4.779)	(8.510)
More than 500 books	78.70***	78.71***	78.71***	74.13***	74.23***	71.96***	72.12***	70.75***
	(6.742)	(6.670)	(3.924)	(8.190)	(6.812)	(4.334)	(5.369)	(7.501)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-291.1***	-292.1***	-292.1***	-371.3***	-290.3***	-217.2***	-148.6**	-100.5
	(56.82)	(56.68)	(38.94)	(63.26)	(44.82)	(43.20)	(58.88)	(73.99)
Average R-squared	0.1450	0.1461	0.1461					
Observations	13098	13098	13098	13098	13098	13098	13098	13098

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

Table A4.4. OLS estimation of the EPF in Reading using mean method (continues)

				Quantile regression results				
	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8)q90
Female	20.64***	20.62***	20.62***	32.61***	27.92***	22.92***	19.08***	16.49***
	(1.841)	(1.825)	(1.256)	(2.348)	(2.134)	(1.763)	(1.653)	(2.117)
Age	11.89***	12.01***	12.01***	13.85***	10.84***	11.54***	11.27***	10.80***
	(3.886)	(3.892)	(2.681)	(5.148)	(2.531)	(2.927)	(2.872)	(3.547)
Immigrant 1st generation	-19.51***	-19.57***	-19.57***	-22.93***	-19.69***	-16.49***	-10.95**	-4.525
	(4.823)	(4.827)	(3.177)	(6.795)	(4.695)	(4.052)	(4.512)	(6.150)
Immigrant 2nd generation	1.536	2.550	2.550	3.819	0.653	-1.292	-0.867	0.969
	(10.13)	(10.12)	(6.609)	(16.39)	(10.42)	(9.852)	(12.77)	(16.62)
Language	-2.214	-1.970	-1.970	-8.831**	-7.097**	-3.992	-4.095	-3.282
	(2.774)	(2.798)	(1.858)	(4.171)	(3.289)	(3.280)	(2.847)	(2.852)
Single parent family	-4.782	-4.666	-4.666**	-9.757*	-4.010	-1.721	-1.726	0.0392
	(3.285)	(3.290)	(1.982)	(4.991)	(3.738)	(2.743)	(2.884)	(2.799)
Mixed family	-26.63*	-26.14*	-26.14***	-30.85*	-26.19*	-15.99	-12.49	-11.12
	(15.15)	(14.87)	(8.881)	(17.95)	(14.24)	(10.55)	(13.18)	(15.26)
Education mother	0.278	0.531	0.531	1.048	0.774*	0.655	0.468	0.254
	(0.352)	(0.551)	(0.399)	(0.687)	(0.458)	(0.469)	(0.398)	(0.496)
Education father	-0.679*	-0.404	-0.404	-0.127	-0.166	-0.182	-0.0859	-0.247
	(0.366)	(0.545)	(0.360)	(0.663)	(0.548)	(0.441)	(0.490)	(0.649)
ISEI mother	0.194**	0.178**	0.178***	0.0929	0.121	0.132*	0.110	0.125
	(0.0783)	(0.0868)	(0.0586)	(0.106)	(0.0771)	(0.0683)	(0.0668)	(0.0926)
ISEI father	0.179**	0.119	0.119*	0.222*	0.216**	0.149**	0.136	0.140
	(0.0832)	(0.102)	(0.0643)	(0.129)	(0.0887)	(0.0746)	(0.0871)	(0.0900)
Overeducation mother		-0.491	-0.491	-7.222	-7.599**	-8.639***	-5.764**	-4.530
		(4.123)	(2.989)	(4.495)	(3.281)	(3.129)	(2.772)	(3.915)
Overeducation father		-6.745	-6.745***	-8.835*	-8.905***	-6.626**	-3.892	-2.237
		(4.322)	(2.553)	(4.686)	(3.397)	(2.867)	(2.755)	(3.830)
Undereducation mother		3.092	3.092	2.119	2.732	3.418	4.868	5.035
		(3.786)	(2.716)	(6.376)	(3.546)	(3.916)	(3.257)	(3.929)
Undereducation father		-1.845	-1.845	-1.900	0.0279	-1.358	-0.591	-0.183
		(3.729)	(2.647)	(5.615)	(3.794)	(3.514)	(3.665)	(4.390)

Table A4.4. (continuation): OLS estimation of the EPF in Reading using mean method

				Quantile regression results				
	(1)	(2)	(3)	(4) q10	(5) q25	(6) q50	(7) q75	(8)q90
From 11 to 25 books	16.63***	16.50***	16.50***	18.61***	19.73***	19.97***	14.83**	16.65***
	(6.177)	(6.132)	(3.865)	(7.102)	(5.325)	(7.098)	(6.122)	(6.159)
From 26 to 100 books	42.43***	42.24***	42.24***	46.73***	44.41***	42.15***	37.28***	38.42***
	(6.161)	(6.101)	(3.195)	(6.324)	(4.967)	(6.178)	(5.683)	(5.932)
From 101 to 200 books	58.48***	58.27***	58.27***	60.32***	57.83***	57.10***	52.11***	52.67***
	(5.818)	(5.746)	(3.509)	(6.787)	(5.060)	(6.288)	(5.923)	(6.092)
From 201 to 500 books	67.71***	67.49***	67.49***	72.83***	69.45***	67.37***	62.62***	64.11***
	(6.294)	(6.224)	(3.801)	(7.246)	(5.344)	(7.044)	(6.436)	(6.219)
More than 500 books	71.36***	71.23***	71.23***	70.14***	69.54***	71.32***	66.30***	67.77***
	(6.518)	(6.480)	(4.306)	(7.685)	(5.400)	(7.167)	(6.652)	(6.539)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-256.4***	-260.1***	-260.1***	-394.4***	-296.7***	-253.7***	-196.6***	-145.3**
	(62.02)	(61.71)	(42.65)	(80.01)	(40.51)	(48.18)	(45.84)	(57.14)
Average R-squared	0.1471	0.1482	0.1482					
Observations	13098	13098	13098	13098	13098	13098	13098	13098

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

Table A4.5. OLS estimation of the EPF

	Mathematics	Sciences	Reading
Female	-21.03***	-11.25***	25.89***
	(2.454)	(2.443)	(2.112)
Age	8.383*	15.77***	13.00***
	(4.347)	(4.029)	(4.316)
Immigrant 1st generation	-24.00***	-24.70***	-23.05***
	(4.591)	(5.141)	(4.371)
Immigrant 2nd generation	-12.80	-6.790	-5.674
	(11.12)	(9.318)	(9.901)
Language	-3.715	-4.597	-6.462*
	(3.580)	(4.945)	(3.681)
Single parent family	-12.21***	-7.322**	-7.526**
	(3.861)	(3.723)	(3.585)
Mixed family	-40.50***	-30.20***	-31.70***
	(10.64)	(10.99)	(11.29)
Education mother	0.634*	1.052***	0.998***
	(0.335)	(0.379)	(0.364)
Education father	0.227	0.0480	0.139
	(0.372)	(0.372)	(0.358)
ISEI mother	0.449***	0.532***	0.458***
	(0.0808)	(0.0817)	(0.0772)
ISEI father	0.494***	0.377***	0.534***
	(0.0791)	(0.0848)	(0.0769)
Mother working	5.097	-0.326	-1.456
	(3.546)	(2.985)	(2.517)
Father working	-2.247	-6.296*	-3.959
	(3.051)	(3.498)	(2.788)
From 11 to 25 books	28.27***	26.46***	25.47***
	(5.584)	(4.831)	(4.674)
From 26 to 100 books	60.80***	55.86***	55.86***
	(5.532)	(5.011)	(4.451)
From 101 to 200 books	85.38***	76.02***	75.12***
	(6.144)	(5.379)	(4.904)
From 201 to 500 books	94.04***	88.73***	83.04***
	(6.155)	(4.864)	(4.633)
More than 500 books	94.51***	91.41***	84.88***
	(6.805)	(5.673)	(5.483)
Regional FE	Yes	Yes	Yes
School FE	Yes	Yes	Yes
Constant	244.9***	142.9**	154.8**
	(68.96)	(65.37)	(69.74)
Average R-Squared	0.275	0.259	0.277
Observations	17370	17370	17370

Note: Standard errors in brackets *** p<0.01. ** p<0.05. * p<0.1.

Chapter 5: Concluding remarks

This dissertation has examined several aspects related to the overeducation phenomenon in Spain. On the one hand, chapters 2 and 3 were related to the fact that overeducated workers have a wage loss compared to properly educated workers with the same level of education. In particular, chapter 2 identified a way to reduce this wage penalty, and chapter 3 tested an assumption to explain this wage penalty. On the other hand, chapter 4 identified a new group that could be suffering negative consequences of overeducation from an intergenerational perspective. This last section summarises the main findings and conclusions from these three empirical analyses as well as their limitations. Finally, it also explaines their policy implications and the future research plan.

The analysis of chapter 2 started from the fact that overeducated workers tend to get more training than their well-educated colleagues at work (Verhaest and Omey, 2006). It analysed whether overeducated workers obtain a higher return on this training – specifically, non-education training activities – than the rest of workers. If it is so, overeducated workers could overcome part of the wage penalty derived from their education-occupation mismatch. The results showed that non-formal education activities have a positive effect on wages, but only overeducated workers who have undergone non-formal education activities receive a wage premium. It seems that this type of training provides overeducated workers with new abilities that permit them to reduce the wage penalisation derived from the mismatch between their level of education and occupation.

The aim of chapter 3 was to test a supported theory based on the existence of individuals' skill heterogeneity (Green and McIntosh, 2007) to explain the wage penalty associated with overeducation. From such a perspective, the wage penalty associated with overeducation is due to the huge variation of skills between workers with the same level of education. Then, overeducated workers would not suffer a wage penalty. In fact, they would earn lower wages as a result of their lower skills. Our hypothesis was that the wage penalty associated with overeducation could be explained by lower skill levels. As a consequence, overeducated workers may not be suffering a wage penalty in Spain, but their earnings are determined by their skill level. Our results showed that individuals' skill heterogeneity explains only 18% of the effect of educational mismatch on wages in Spain. The wage penalty still remains for those overeducated workers who are not less skilled than properly matched workers.

Next, chapter 4 aimed to identify a relationship between the situation of overeducation of parents and the educational performance of their children. As Haveman and Wolfe (1995) highlight, children of highly educated parents tend to perform better than children of less educated parents. One possible explanation for the positive relationship between parents' human capital and students' performance is based on children's perceptions about the importance of education. In this sense, students whose parents have a high level of education and good jobs might be more aware of the value of education and, consequently, have higher motivation and perform better than other students. Under this point of view, our hypothesis was that the existence of parents' job-education mismatch can modify the students' perception about the importance of education and, consequently, have an effect on their performance at school. In particular, we analysed whether there is a relationship between parents' educational mismatch and the educational performance of their children, and we checked whether it is similar across the performance distribution or, by contrast, whether there are differences between students at the top and at the bottom of the performance distribution. The results showed that students whose parents are overeducated have a penalty in their academic achievement in all three subjects analysed, this effect being stronger for students with lower educational outcomes. So, the results seemed to confirm our hypothesis.

Unfortunately, this dissertation is not without its limitations. The wage model estimated in chapter 2 could have an endogeneity problem related to the omitted variable – individual ability – that is contained in the error term. Then, education is probably correlated with the error term and with wages (see for a discussion Leuven and Oosterbeek, 2011). A growing number of studies have indicated that the omission of ability overstates the wage penalty for being overeducated (Chevalier, 2003; Frenette, 2004; Chevalier and Lindley, 2009; Tsai, 2010; Pecoraro, 2014). However, we are not able to overcome this limitation, because AES is cross-section data that has rich information on training activities but not on workers' ability. Conversely, chapter 3 estimated a wage model including explicitly individual skills. It is worth noticing that the effect of overeducation on wages remains negative and statistically significant, whereby the omission of the ability variable in the model of chapter 2 could not be a big problem in trusting the results. Finally, chapter 4 was based on the estimation of an education function production to find a relationship between overeducated parents and their children's performance in school. It was found that this relationship is negative and statistically significant, but we cannot ensure the direction of the effect. In other words, we could have an endogeneity problem related to the simultaneity. For example, the bad match of parents in the labour market could negatively affect their children. However, another possibility is that the bad performance of children causes their parents to decide to get a worse job that allows them to work fewer hours in order to spend more time with their children. However, we have not found a proper instrumental variable to overcome this problem, since PISA data have little information about parents. Moreover, our hypothesis considered that the transmission channel of this relationship is based on students' perceptions, but we were not able to test it. Thus, findings in chapter 4 should be interpreted cautiously.

As for the public policies required to reduce overeducation in Spain, we think that there is much work to do. First of all, it is worth highlighting that overeducation is not only a phenomenon that has an effect at an individual level, but entails a waste of public resources given that high education is highly subsidised. Thus overeducation should be addressed by governments in order to reduce its incidence and consequences. For instance, public policies addressed to incentivise the creation of innovative companies should be a priority in the agenda. Otherwise, Spain will lose an important portion of the high-skilled workers it produces, and it will not be able to attract this type of worker in the future. From the educational point of view, it would be interesting to implement the following ideas. First, it would be interesting to know what companies require from their high-skilled workers in order to adapt part of the curricula of the university programs. Second, educational institutions should also give all the information about the employability of each type of education to students before they start a specialised course. And they should also encourage students in entrepreneurship. Self-employment could be a way to overcome the lack of demand for specific workers.

Furthermore, other policy implications could be identified from the findings explained previously. From the findings of chapter 3, as part of the effect of overeducation on wages is due to a lack of competence or skills of overeducated workers, educational policy makers should focus on defining the level of competence or skills that should be acquired at each level of education. Indeed, skills should be evaluated at educational institutions in the same way as education. On the other hand, chapter 2 pointed out the role of training – in particular, non-formal education activities – for reducing the wage loss of overeducated workers. So, it could be effective to design policies to encourage not only the less educated population to get training, but also the rest of workers. The analysis presented in chapter 4 identified children as a new collective that could be affected by their overeducated parents. This finding illustrates the importance of the quality of educational systems that guarantee an appropriate performance of students regardless of their family situation.

Last, there are questions related to these three studies that are left for further research. For example, the study presented in chapter 2 used data from the Adult education Survey from 2007, and a new wave for 2012 is already available. It could be interesting to analyse whether the value of training has changed over time and whether overeducated workers are still the ones who get a wage benefit from these training activities using both waves of AES.

The same strategy could be interesting in the analysis presented in chapter 4. It considered PISA data from 2009 to verify the relationship between parents' overeducation and educational outcomes of their children. In this case, there is also a new wave of the database available (PISA 2012). It could also be interesting to analyse that relationship before and after the economic crisis to give more support to our hypothesis about the transmission channel of this relationship. We consider that the lower educational achievement of children whose parents are overeducated could be related to their perception about the value of education. Thus, if their high-educated parents end up in a job that requires less education, they may underestimate the value of education.

Finally, there are many interesting analyses related to the one presented in chapter 3. PIAAC data is only available since 2013 and provides detailed information about individuals' skills level, skills used at home and skills used at work. In this sense, evidence in chapter 3 showed that only the 18% of the wage penalty suffered by overeducated workers was explained by their skills level. The interesting question that follows is whether high-skilled, overeducated workers are fully utilising their skills at their jobs or not. This analysis will allow us to determine whether companies take advantage of the skill level of overeducated workers or, by contrast, whether they perform the same job better than their well-educated colleagues. Furthermore, we believe that it could be interesting to compare the results obtained for Spain with those for other countries. All three databases are indeed available for many other countries, so these would allow us to perform a comparative analysis of the differences between countries.

References

Chevalier, A. (2003) "Measuring Overeducation", *Economica*, 70 (3), 509-531.

Chevalier, A. and Lindley, J. (2009) "Overeducation and the skills of UK graduates", *Journal of the Royal Statistical Society: Series A*, 172, 307-337.

Frenette, M. (2004) "The overqualified Canadian graduate: the role of the academic program in the incidence, persistence, and economic returns to overqualification", *Economics of Education Review*, 23, 29-45.

Green, F. and McIntosh, S. (2007) "Is There a Genuine Under-utilisation of Skills Amongst the Over-qualified?", *Applied Economics*, 39(4), 427-439.

Haveman, R. and Wolfe, B. (1995) "The determinants of children's attainment: A review of methods and findings", *Journal of Economics Literature*, 33(4), 1829-1878.

Leuven, E. and Oosterbeek, H. (2011) "Overeducation and Mismatch in the Labor Market", in: E. Hanushek, S. Machin, L. Woessman (Eds.), *Handbook of the Economics of Education*, 4, 283-326.

Pecoraro, M. (2014) "Is there still a wage penalty for being overeducated but well-matched in skills?" A panel data analysis of a Swiss graduate cohort" *Labour*, forthcoming.

Tsai, Y. (2010) "Returns to overeducation: A longitudinal analysis of the U.S. labor market", *Economics of Education Review*, 29, 606-617.

Verhaest, D. and Omey, E. (2006) "The impact of overeducation and its measurement", *Social Indicators Research*, 77, 419-448.