Overeducation, skills and wage penalty: Evidence for Spain using PIAAC data

Sandra Nieto

AQR-IREA Research Group, Universitat de Barcelona

snieto@ub.edu

(Preliminary version, February 2014)

The literature on educational mismatches finds that overeducated workers suffer a wage penalty compared with properly educated workers with the same level of education. Recent literature also suggests that individuals' skill heterogeneity could explain that wage's differences between overeducated and properly matched workers. The hypothesis is that overeducated workers earn less due to their lower competences and skills in relative terms. However, that hypothesis has been rarely tested due to data limitations on individual's skills. The aim of this paper is to test the individual's skill heterogeneity theory in Spain using microdata from PIAAC, since it is one of the developed countries supporting the highest overeducation rates and where its adult population holds the lowest level of skills among a set of developed countries. Our hypothesis is that the wage penalty of overeducation in Spain is explained by the lower skill level of overeducated workers. The obtained evidence confirms this hypothesis but only to a certain extent as skills only explain partially the wage penalty of overeducation.

JEL Codes: J31, I21, C13

Keywords: overeducation, individual's skill heterogeneity, wages.

Acknowledgements: I thank Raul Ramos for his dedication and guidance and Elena Costas-Perez for her helpful comments. I also acknowledge the support from AQR-IREA Research Group.

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 job characteristics. 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. 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 individual's 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 be suffering a wage penalty. In fact, they would earn lower wages as a result of their lower skills. If this hypothesis holds, wage penalty will disappear once individuals' skill level is included in the analysis. However, most of the literature does not explicitly test that hypothesis due to data limitations regarding individuals' skill levels.

In this paper we take advantage of the recently available database of the OECD Programme for the International Assessment of Adult Competencies (PIAAC) since 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. It is one of the developed countries supporting one of the largest percentage 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[1] has one of the lowest levels of proficiency in literacy and numeracy skills (OECD, 2013a).

Therefore, the specific aims of the paper 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 individual's 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 could not be suffering a wage penalty in Spain, but their earnings are determined by their skill level.

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 paper is structured as follows. First, section 2 provides a literature review on the analysis of skills in educational mismatch. Section 3 introduces the PIAAC data and explains how educational and skills mismatch is measured. Section 4 shows the relationship between overeducation and overskilling. Next, Section 5 quantifies the wage penalty of overducation and the impact of skills using different specifications. Last, Section 6 concludes with some final remarks.

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

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 often regarded are the assignment model, as well as individual's skills heterogeneity.

The assignment theory (Sattinger, 1993) makes the assumption that the human capital' returns depend on both, the workers' human capital and the match between the worker and the job. From such a perspective, workers' productivity is limited by their job characteristics. 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 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 McGuiness, 2013; Mavromaras et al. 2013). It is found that overeducation and overskilling have both a negative and statistically significant effect on earnings within the same level of education, being the overeducation effect much higher than the overskilling effect. This result points out that wage penalization associated with overeducation is not explained by under-utilization or waste of workers' skills, whereby the assignment theory is not supported by results. They may suggest the existence of heterogeneity of workers' skills. However, they do not explicitly test that 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 their job characteristics. Therefore, the observed wage differences among overeducated and undereducated workers compared to well-matched workers with the same educational level might just reflect individual differences in human capital within educational levels. In other words, overeducated workers are less productive because they have less human capital, not because their job imposes limitations in their productivity.

As it has been mentioned before, data availability on workers' skill levels is very limited, whereby different approaches have been considered in empirical analysis trying to control by individual skills heterogeneity in the wage equation's estimation.

One of the approaches 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 disappear when it is estimated by fixed effects; suggesting that (part of) the effect of educational mismatch is caused by unobserved individual ability.

Instead of using a longitudinal framework, Chevalier (2003) analyses cross-sectional data. He creates a proxy of workers' unobserved productivity taking the difference between the estimated and the observed earnings in their first job. 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 to analogous results. They construct the measure of unobserved ability as the residual from a first-job earnings equation, capturing all individual's observed characteristics including those job characteristics that affect wages. Those 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 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 using a direct measure of skills. Korpi and Tåhlin (2009) also include explicit indicators of ability in an analysis for Sweden using panel data, and the effect of overeducation on wages still remains statistically significant.

The recent study of Levels et al. (2013) also includes individual's 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 to 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; Nieto and Ramos, 2013; among others),to our knowledge, no longer analysis has tested the role of individuals' skill level on educational mismatch literature focusing in the Spanish case.

3. Data sources and variable definition

3.1. PIAAC database

The Programme for the International Assessment of Adult Competencies (PIAAC) is a survey which has been conducted by 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's countries) between 2011 and 2012.

Participation in the problem-solving domain was optional, and Spain (and other countries) did not participate in it. In consequence, the competences we analyse are related to literacy and numeracy. Specifically, both 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), that are represented in 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 those scores using answers from the test and background questionnaire of similar individuals. It generates a distribution of values for each individual and their associated probabilities whereby ten plausible values are randomly obtained for each individual. This method prevents the 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[2].

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 the 10 plausible values test scores in literacy. The second one is composed by other personal, job related and regional variables which are included in the model as controls[3]. These variables are: gender, age, nationality, type of contract (full-time/part-time), contact term (temporary/permanent), sector

---

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

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

(public/private), economic activity (industry, agriculture, construction, services, non-sale services) and 17 regions.

The initial Spanish sample is composed by 6.055 observations. We restrict the sample to employed workers who are not in education at the same time. We drop of the analysis armed forces workers, and those that did not give some of the information we need to perform the analysis. The final sample is 1.928 observations. Table 1 of Annex shows the descriptive analysis of the variables previously defined.

3.2 Measuring educational and skill mismatches

There are different methods to measure educational mismatch: the objective, the subjective or workers' self-assessment and the statistical or realized matches. All of them have advantages and drawbacks, whereby using either measure method finally depends on the availability of the data (see Hartog, 2000; and Leuven and Oosterbeek, 2011; for a review).

PIAAC data allows us measuring required schooling by 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. PIAAC questionnaire contains specifically 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 comparing workers' answers about required education and attained education. Then, 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 those required by their jobs.

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 determining 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 when their education differs the mode in their occupation.

Table 1 shows the impact of educational mismatch in Spain using the self-assessment method. About half of workers in Spain have a proper match between their education and occupation. From the remaining workers, PIAAC data points out that overeducation is affecting to 35.63%, being 3.8 the average of surplus' years of education. And, on the other hand, undereducation concerns the rest 15% and their average of deficit's years of education is 3.1.[4] [5]

Table 1: Educational mismatch

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

Source: Own elaboration using PIAAC data. Individual sample weights have been considered.

The percentages obtained by the statistical method are shown in Table A.2 of Annex. It is worth to notice that different measurement methods use to report different percentages of educational mismatch although it is considered the same country and the same database. However, the impact on wages uses to be consistent regardless the measurement method considered (Hartog, 2000). We perform the main analysis measuring educational mismatch by the self-assessment method and we repeat it using the statistical method as a robustness check.

As regards 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 between 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,

[4] 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.

[5] Similar incidence of educational mismatch in Spain has been found by Murillo et.al. (2012)

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 well match between their skills and those required by their jobs. Moreover, overskilling affects to 21.4% of workers whereas 6.5% are underskilled.

4. Are overeducated also overskilled?

As it has been explained in the previous sections, individuals' skill heterogeneity is one of the explanations to 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 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 out differences that could suggest the existence of individuals' skill heterogeneity.

PIAAC data for Spain shows that all workers have a higher probability of being well-skilled, regardless 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 surprisingly that only the 7.5% of undereducated workers are also underskilled, and the 20% have an excess of skills. However, 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[6].

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

Table 2: Distribution of undereducated, well-matched and overeducated workers by their skill (mis)match in literacy (in %)

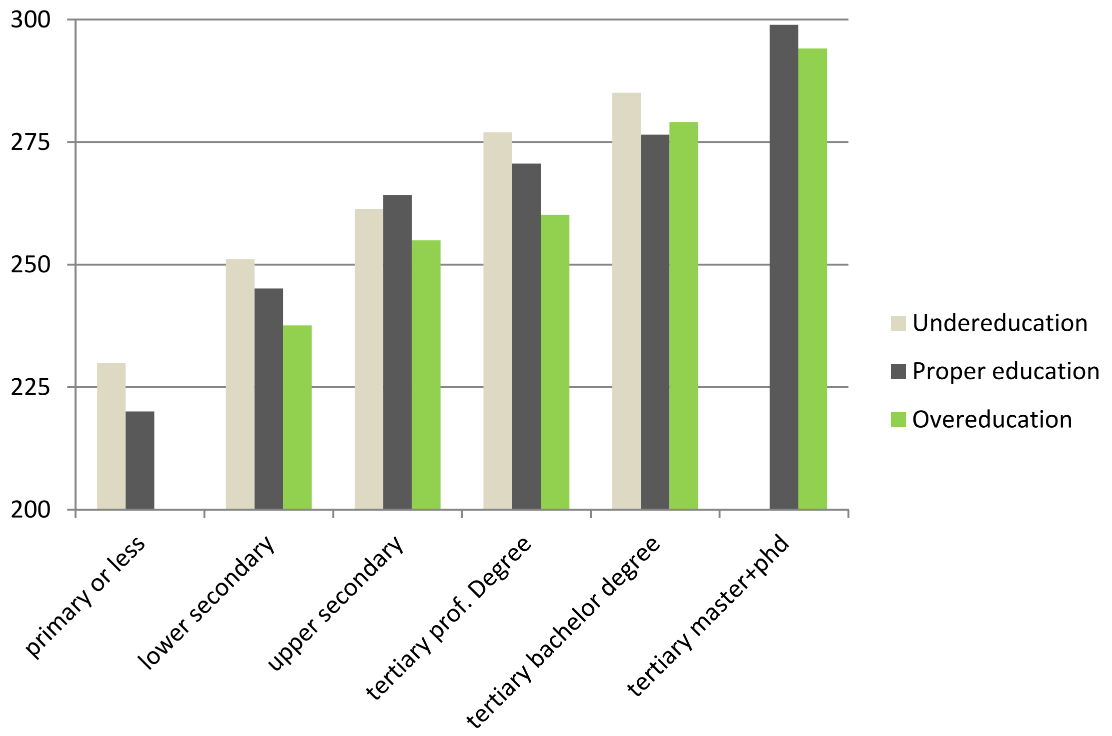
	Underskilling	Proper skills	Overskilling	Total
Undereducation	7.48	72.24	20.28	100.00
Proper education	6.52	73.18	20.29	100.00
Overeducation	6.02	70.62	23.36	100.00

Source: Own elaboration using PIAAC data. Individual sample weights have been considered.

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

Figure 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 proper educated workers with the same educational level (except for upper secondary education).

Figure 1: Average skills levels of workers by educational level.



Source: Own elaboration using PIAAC data. Individual sample weights considered.

Thus, data show skill heterogeneity between workers with the same level of educational. That fact could explain the wage's differences between workers depending on their education-occupation match. It is empirically tested in the following sections 4 and 5.

5. Educational mismatch, skills and wages

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 by 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 as skills or ability, whereby we also include the individuals' skills in all three models.

Specifically, the traditional wage equation is defined as:

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

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 which also includes other human capital's variables as experience, experience squared and a dummy variable which takes de value 1 if the worker has participated in some non-formal education activity during the last 12 months previous 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:

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

Where $skills_i$ is a continuous variable measured by scores in a 500-points 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-Undereducated) specification created by Duncan and Hoffman (1981). This specification split 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, it is determined:

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

The ORU equation is then defined as:

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

The other variables' definitions are the same as in specification (1a). The interpretation of the coefficients associated to over- and undereducation are compared with well-matched workers in the same job. 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 individual's skills:

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

The variable $skills_i$ is defined as in equation (1b). If individual's skills heterogeneity completely explains the wages' effects of educational mismatch, we should get that $\beta_1 = \beta_3 = 0$. If it is so, workers' remuneration composed by their education and skills would be determined by required education and their individual's skill level.

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

The V&V equation is defined as:

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

Where OE is a dummy variable which 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 to 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:

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

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

5. 2. Results

In line with similar studies (see, for instance, Dolton and Vignoles, 2000; and Di Prieto and Urwin, 2006), we control in all the previous specifications for a possible problem of sample selection bias estimating by Heckman two step specification (Heckman, 1979). That procedure takes into account the possibility that employed workers may not be a random subsample of the sample we are finally considering. The first step estimates the probability of being employed using a probit equation[7] (see the results in Table A.3). Then, the probit

[7] 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.

estimation is used to construct a selection bias control factor, which is included as an explanatory variable in the wage equation[8].

As we focus on the analysis of the variables related to human capital, we only comment the results of those variables in the main test. However, it is worthy to notice that the coefficients associated to control variables are similar of those in the previous literature. Furthermore, it is found that the lambda coefficient is positive and statistically significant for all specifications. Hence, the omission of the information about the probability of being employed in the wage analysis would imply a bias in the results.

Table 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 equation (3a) and (3b).

As concerns the traditional Mincer's models, it is shown that the return of the variables related to human capital are similar of what previous literature finds (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. And 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 a statistical 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-points 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

[8] 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.

undereducation. In particular, overeducation workers obtain for each surplus year of education a 3% of 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 it is specified in equation (2b). Skills have a positive a statistically significant effect on wages, but the effects of educational mismatch still remain statistically significant. So, 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 by 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, whereby our initial hypothesis about the Spanish case is not supported by data. Specifically, skills only explain the 18% of the wage's effect of overeducation and the 14% of the effect of undereducation on wages. The obtained results are in line with what the analysis of Levels et al. (2013) finds for a set of OECD countries.

Table 3. Estimated earnings functions

	Mincer		ORU		V&V	
VARIABLES	(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.

Finally, the results from V&V specifications defined in equations (3a) and (3b) are shown in the last columns of table 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 to overeducated workers reduces from 17.1% to 16.3%. On the other hand, the premium of undereducated workers falls from 13.1% to 12.4%. As it is found in the results from the ORU specifications, those results neither seems to support the heterogeneity skills theory, since the effects of overeducation and undereducation still remains once it is controlled for skills.

To sum up, we obtain that individuals' skills are important to determine individuals' wages as the 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 the 18% of their lower wages compared to well-matched workers with the same level of education.

5.3. Robustness checks

PIAAC data allows us doing some robustness check to validate the previous results.

First, literature shows that the incidence of both overeducation and undereducation could be different depending on the measurement's method applied. However, the effects on wages are quite consistent regardless the measurement method. Besides the self-assessment method, 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 A.4. of Annex). Specifically, individuals' skills only explain the 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, whereby we decided not include both together. We also estimate the ORU specifications including numeracy skills instead of literacy skills (Table A.5. of Annex). The results show that the wage penalty of overeducated workers is reduced but it still remains once skills are included. Specifically, individual's skills heterogeneity explains the 22% of that wage penalty.

6. Final remarks

The main objective of this paper is to analyse whether individual's skill heterogeneity explains wage penalty of overeducated in Spain. Our hypothesis is that the wage penalty associated to overeducation could be explained by the low skill level of overeducated workers, since Spain holds the lowest level of skill between its population. In consequence, overeducated workers could not be suffering a wage penalty in Spain, otherwise their earnings are determining 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 to the previous results. On the one hand, as part of the effect of overeducation on wages is due to a lack of competences or skills of overeducated workers, educational policy makers should focus on define the level of competences or skills should be acquired in each level of education. Indeed, skills should be evaluating at educational institutions as the same way that 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 encouraging students to entrepreneurship. Self-employment could be a way to overcome the lack of demand of specific workers. And, finally, Spanish government should do 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.

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." IZA DP No 5523.

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.

Nieto, S. and Ramos, R. (2013) "Non-formal education, overeducation and wages", Revista de Economía Aplicada, 61(XXI), 5-28.

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.

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, in press. http://dx.doi.org/10.1080/09645292.2013.846297

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:

Table A.1. Descriptive statistics

Variable	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

Source: PIAAC. Individual sample weights considered. Number of observations 1928.

Table A.2. Educational mismatch using the statistical method.

Variable	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

Source: Own elaboration using PIAAC data. Individual sample weights considered.

Table A.3. Heckman's specification first step. Determinants of being employed.

VARIABLES	(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

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

	Mode		Mean	
VARIABLES	(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 A.5. Estimated earnings ORU functions using numeracy skills

VARIABLES	(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

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.