Caroleo, Floro Ernesto; Pastore, Francesco

**Working Paper**

# Overeducation at a glance. Determinants and wage effects of the educational mismatch based on AlmaLaurea data

GLO Discussion Paper, No. 15

**Provided in Cooperation with:**
Global Labor Organization (GLO)

*Suggested Citation:* Caroleo, Floro Ernesto; Pastore, Francesco (2017) : Overeducation at a glance. Determinants and wage effects of the educational mismatch based on AlmaLaurea data, GLO Discussion Paper, No. 15, Global Labor Organization (GLO), Maastricht

This Version is available at:
https://hdl.handle.net/10419/155277

# Overeducation at a glance.
# Determinants and wage effects of the educational mismatch based on AlmaLaurea data

**Floro Ernesto Caroleo[¥] e Francesco Pastore[§]**

## Abstract [#]

This essay delivers two main innovations with respect to the existing literature. First, and foremost, by extending the work of Nicaise (2010) relative to the reservation wage to the case of overeducation, we propose a statistical test to discriminate between alternative theoretical interpretations of the determinants of overeducation through the Heckman sample selection procedure. Second, the essay provides the first available economic analysis of the consequences of the educational mismatch in Italy as based on AlmaLaurea data, the largest and richest data bank available in the country. The data includes a large number of university graduates enrolled in a given year before the Bologna reform and asks a large number of questions allowing us measuring among others the quality of education from high school. This wealth of information is a condition to provide the most comprehensive, accurate and reliable assessment of overeducation in the country. The educational mismatch 5 years from graduation is relatively high – at 11.4% and 8% for overeducation and overskilling, respectively – by EU standards. *Ceteris paribus* the parents of the mismatched have lower educational levels according to school tracking. Most humanities and social sciences degrees but also geology, biology and psychology are associated with both types of mismatch. The quality of education also correlates to the educational mismatch. We find a non-conditional wage penalty associated to overeducation and overskilling of 20% and 16% and a conditional

[¥] Floro Ernesto Caroleo is professor of Economic Policy at the University of Naples "Parthenope" and a research fellow of the IZA (Institute for the Study of Labor) of Bonn. Email: caroleo@uniparthenope.it.

[§] Corresponding author: Francesco Pastore is Francesco Pastore is associate professor of Political Economy at Seconda Università di Napoli and a research fellow of the IZA (Institute for the Study of Labor) of Bonn. Email: francesco.pastore@unina2.it.

[#] Acknowledgements. Most part of this research was conceived during several periods of stay of Francesco Pastore at the AlmaLaurea headquarters in Bologna. We are grateful to Andrea Cammelli, Angelo Di Francia and Silvia Ghiselli not only for providing us with AlmaLaurea data, but also for detailed and helpful suggestions on data analysis. Previous versions of this paper have been presented at: AlmaLaurea Conference, Bologna (2011), IAB of Nuremberg (2011), XXVI AIEL Conference, Catholic University of Milan (September 2011), University of Naples "Parthenope" (2011), University of Naples "Federico II" (2011), Second University of Naples (2011; 2012), University of Pescara (2012); University of Salerno (2013). We thank all seminar participants and particularly Lilia Costabile, Giuseppe Croce, Hans Dietrich, Francesco Ferrante, Claudio Lucifora, Nadia Netti, Patrizia Ordine, Marco Pecoraro, Claudia Pigini, Giuseppe Rose, Dario Sciulli, Francesca Sgobbi and Peter Sloane for valuable comments on earlier versions of this paper. This notwithstanding, the authors are solely responsible for the opinions expressed in this paper.

one of about 12% and 7%, respectively. The Heckman sample selection model returns a slightly higher sample selection corrected wage penalty, supporting not only the job competition and job assignment models, but also the human capital model. Other concurrent statistical tests point to the difficulty that the educational system faces in providing work-related skills to graduates.

## Keywords

University-to-work transition; educational mismatch; sample selection bias; AlmaLaurea; Italy.

## JEL classification

C25; C26; C33; I2; J13; J24.

**List of Abbreviations**

| | |
|---|---|
| CEDEFOP | European Centre for the Development of Vocational Training |
| EU | European Union |
| GSOEP | German Socio-Economic Panel |
| IALS | International Adult Literacy Survey |
| IR | Interval Regression |
| ISFOL-Plus | Istituto per lo sviluppo della formazione professionale dei lavoratori (En. Tr.: Institute for the vocational training of workers) |
| ISTAT | Italian Institute of Statstics |
| IV | Instrumental Variables; |
| NUTS | Nomenclature of Territorial Statistical Units (*Nomenclature des unités territoriales statistiques*); |
| LSCU | Laurea Specialistica a Ciclo Unico |
| ML | Maximum Likelihood |
| OECD | Organization for Economic Cooperation and Development |
| OLS | Ordinary Least Squares |
| ORU | Over-, Required and Undereducation |
| PIAAC | Program for the International Assessment of Adult Competencies |
| REFLEX | Flexible Professional in the Knowledge Society |
| ROC curve | Receiver Operator Characteristic Curve |
| UK | United Kingdom |
| USA | United States of America |

# Introduction

Despite the recent dramatic rise in the supply of graduates in the youngest cohorts (25-34 years old) from just over 10% to about 24% in the last two decades, both the supply of and the demand for skills remain lower in Italy than the EU average (OECD, various years). However, there is growing concern that the increased supply of skills in a period of stagnation of productivity growth might be causing a further increase in the proportion of young people forced to work in positions that are below their educational level or that underuse their skills. In a recent study by Cedefop for the European Commission (2013: Chart 7), with over 20% of overqualified graduates, Italy is in 6$^{th}$ position within the EU27, despite the fact of being the last in terms of tertiary education attainment. The case of Italy – increasing supply of skills in a period of persistent economic crisis – is indicative of a number of other EU countries, especially those located in the Southern peripheral areas, where the supply of skills is on the rise, despite the insufficient demand for skills. The lack of systematic and homogeneous information does not allow us to fully assess the extent to which the educational mismatch has changed over recent years and as a consequence of the economic crisis, but it allows us delivering the most vivid, clear and reliable assessment of the extent, determinants and consequences of the educational mismatch in the country.

To do so, we provide the first available study based on the AlmaLaurea database. As further discussed in the data section, this is by far the best available source of statistical information in the country, as it covers a large number of Italian universities. In addition, AlmaLaurea contains a wealth of information about the quality of education and the overall process of school-to-work transition, which is unavailable in other data banks. This allows us to elaborate not only a very detailed and comprehensive analysis of the determinants and consequences of the mismatch, but also a new interpretation of it outlining the way in which the university-to-work transition is organized. In fact, one of the main points of this paper is that the relatively high share of mismatched graduates typical of Italy could be due not only to an excess supply of (or a low demand for) graduates, but also (or rather) to the lack of work-related skills of graduates and to the difficulty of the school-to-work transition system in providing skills which can be used in business.

The new interpretation we provide is likely to apply not only to the Italian case but to all southern European countries and, more generally, to all countries where the educational system is sequential (rather than dual), hence assuming its mission to be solely the provision of general education rather than human capital as a whole (for the USA, see Cappelli, 2015; for a cross-country perspective, Pastore, 2014). The effectiveness of this system in preventing the

educational mismatch relies heavily on the quality of graduate training and early labour market experiences. However, the high degree of labour market rigidity, the recent precariousness of early labour market experiences and the low expenditure in pro-active schemes make early labour market experiences relatively unsuccessful.

Furthermore, our conclusion regarding the Italian case is perfectly in line with some recent interpretations of overeducation not as a departure from, but, conversely, as indirect confirmation of the assumptions and conclusions of the human capital theory (see, in particular, Leuven and Oosterbeek, 2011; Verhaest and Van der Velden, 2013).The second and also the most important contribution of this paper consists of proposing a new interpretation of the Heckman procedure to control for sample selection bias used for estimating the wage penalty of overeducation.. In fact, we extend to the case of overeducation the previous analyses by Nicaise (2001) relative to the determinants of the reservation wage. We submit that the Heckman sample selection procedure can be used as a screening device to choose among different theoretical interpretations of overeducation. An increase of the OLS base coefficient would be suggestive of a lower skill level of the non-employed, which is consistent with the assumptions of the job competition, job assignment and human capital model; vice versa, a reduction in the OLS base coefficient would be suggestive of a higher unobserved skill level of the non-employed, which is consistent only with the assumptions of the job search model.

Although the Heckman procedure has already been used in the context of estimating the wage penalty associated to overeducation (Sloane et al., 1999; Dolton and Vignoles, 2000; Cutillo and Di Pietro, 2006), no previous study has brought to the fore, as we do in this paper, the correct interpretation of the key coefficients obtained from the Heckman procedure in terms of the different theories of overeducation. This is a contribution, which is relevant not only for the overeducation literature, but also for the applied statistics and econometrics literature as well. Generally speaking, theoretical and econometric modelling follow parallel lines, but here we build a bridge between them.

Based on our empirical analysis of the AlmaLaurea databank, we find that the job search model should be discarded in favor of the other three models. Further empirical evidence has been also used to detect the preferred one among the three models selected as based on the Heckman procedure. In fact, the job competition and job assignment models assume that the probability to be overeducated is equal across individuals with the same level of human capital; vice versa, according to the human capital model, overeducation arises because of a lower quality of human capital, due, in the Italian case, to the lack of work-related competences. We show that overeducation is indeed associated with a lower quality of human capital, as shown in the analysis on the determinants of overeducation. In turn, this allows us establishing a link

between the analysis of the wage penalty associated to overeducation and that of the determinants of overeducation, which have been considered independent of each other previously.

This essay is structured as follows. Section one provides a short theoretical survey strictly aimed to define the grid of predictions for the subsequent empirical analysis. Section two summarizes the relevant empirical literature on the Italian case. Section three delivers the new interpretation of the Heckman procedure in terms of alternative theoretical explanations of overeducation, as discussed in section one. Section four describes the AlmaLaurea data in detail. Section five presents the main findings of the empirical analysis. After the descriptives, we investigate, firstly, the determinants of overeducation and overskilling by Logit and, then, we estimate the conditional wage penalty using OLS. In the subsequent subsection, we present the results of the Heckit econometric specification of the earnings equation. In the last subsection, we present and debate of several robustness checks. The concluding section also confers on possible policy suggestions to reduce the size of the educational mismatch.

## 1, Theoretical explanations of overeducation

Theories that explain overeducation include different conceptual constructs: a) the human capital model; b) the job competition model; c) the job assignment model; d) the job search model; e) and the career mobility theory (for more detailed surveys, see, Sloane 2003, McGuinnes 2006, Leuven and Oosterbeek 2011). For shortness' sake this section will just shortly summarise these theories with the aim to just formulate testable predicitions for the empirical analysis of the ensuing sections.

Traditionally, overeducation has been considered an exception to the human capital model as it is associated to a mismatch and, therefore, to a market disequilibrium (see, for instance, Büchel et al., 2003). According to this line of reasoning, it should be seen, hence, as a short term phenomenon since a sufficient degree of wage flexibility should restore any imbalances between supply and demand in the graduate labour market unless some persistent, often unobserved, low ability / skill problem affects the permanently overeducated.

However, as Leuven and Oosterbeek (2011), argue, among others, overeducation does not need to represent a breach of the validity of the human capital model: in fact, overeducation could be conceived as a consequence of a lack of the work-related component of human capital, rather than a waste of human capital. Developing this intuition from an empirical point of view is one of the main achievements of this paper. We can reconcile overeducation with the human capital model by resorting to the point made by Gary Becker (1962) himself according to whom

human capital does not include only general education, but also the general component of work experience and the component of work experience that is specifically attained by working for a sufficiently long period of time on a particular type of job. Now, overeducation could be seen as a consequence of a lack of skills that could be only acquired through work experience and this is typical of young people, despite their increasing educational level. In this theoretical framework, overeducation is not a consequence of an excess, but of insufficient human capital or human capital of inferior quality. Accordingly, overeducation could be seen as one of the forms of the lower return to this low quality of human capital in a logical chain which goes from a greater probability of non-employment and, in case of employment, of a greater probability to find jobs of lower quality, which are destined to diploma holders rather than university graduates. This allows us defining a very precise prediction regarding the definition of the estimated coefficients of the sample selection corrected measure of the wage penalty of overeducation, as will be further discussed in the methodological section.

One would expect hence that overeducation be greater where the educational system is of a sequential type, namely where the mission of the educational system is generating general education rather than all-round human capital, as it is instead the case of dual educational systems (for a detailed classification of school-to-work transition regimes, see Pastore, 2014).

Lester C. Thurow (1979) brought to the fore for the first time the job competition model, which helps understanding the *persistence* of overeducation also among the adults. In this case, excess schooling is a consequence of the competition for jobs in presence of rigidity of the demand for highly educated labour that leads graduates to accumulate education in order to reach the best positions in the queue for the job. In this process, the amount of education acquired by workers is in some cases more than the amount actually requested to get or to do a job. The job competition model assumes that education does not necessarily increase productivity, but is more simply and essentially a signal of skills that job candidates transmit to their perspective employers. Again, here, the queue is due to high unemployment and the penalty for some of the university graduates is, first of all, to remain non-employed and, in case of employment, to experience overeducation. In the job competition model, there is no specific reason why some university graduates get the best jobs and others only low quality jobs or no job at all. It is a random process. Simply the labor market generates less graduate jobs than those requested and the most lucky will get the best jobs.

With the job assignment theory, Sattinger (1993) attempted to reconcile the two previous theories of human capital and job competition. Like the job competition model, the assignment model assumes that the graduate jobs available in the economy are limited, which implies that remuneration is job specific and partly independent of the human capital endowment of the

individual; on the other hand, like the human capital theory, it assumes that with their investment in human capital individuals are able to compete for the best jobs and wages are bound to be influenced by the human capital level of individuals. Overeducation arises because wages will neither be entirely related only to attained schooling and other individual attributes, like in the human capital model, nor only to the nature of the job itself, like in the job competition model.

In its standard formulation, the job search theoretical model assumes, instead, that unemployment is largely a voluntary choice. People accept a job offer when it is associated to a wage higher than or equal to their reservation wages. The most skilled graduates have higher reservation wages and wait for a longer time than the least skilled graduates, who tend to choose the first job offer they can get, even if it involves overeducation, because of their lower reservation wage. Albrecht and Vroman (2002), Dolado et al. (2009) and Carroll and Tani (2013) are examples of this stream of the literature.

Overeducation may result also from career mobility theories, which is linked to the old idea of overeducation as a breach of the human capital model (Sicherman and Galor, 1990; Buchel and Mertens, 2000): wages tend to grow over time together with the work experience accumulated by individuals. It would be, therefore, physiological that firms and graduates generate job-worker matches with low earnings in the short run, but good career prospects in the long-run.

Opposite to this last approach is, however, the recent empirical literature which finds that skills under-utilization may also have a scarring effect, similar to unemployment and low pay jobs (Mavromaras et al., 2015; and, on a similar vein, Congregado et al., 2016). Baert et al. (2013) adopt the timing of events approach to identify a dynamic treatment effect of overeducation on future employment chances as foreseen by the career mobility theory using monthly calendar data from a representative sample of Flemish (Belgian) youth who started searching for a job right after leaving formal education. They find that, in fact, overeducation is a trap rather than a stepping stone. And nonetheless, Baert et al. (2014) find that the negative signal of overeducation is still less strong than that of unemployment.

Based on the previous discussion, we can formulate the following testable predictions or hypotheses (TH) for the empirical analysis, so as to form a sort of grid based on which we can select the preferred theoretical interpretation among all those now mentioned:

$TH_0$: according to the human capital, job competition and job assignment models, the non-employed would be overeducated if employed;

$TH_1$: according to the job search model, the non-employed would not be overeducated if employed;

This is a clear hypothesis, which is testable within the Heckamn smple selection framework, as shown in Section 3. In addition, $H_0$ can be further subdivided in two parts to disentangle the three models under $H_0$:

$TH_{00}$:the overeducated possess a human capital of lower quality than average (Human capital model);

$TH_{01}$:the overeducated possess a human capital of the same quality as average (job competition model).

In fact, the case of the job assignment model is partly compatible with both $H_{00}$ and $H_{01}$. Last, but not least, also $H_{00}$ can be further subdivided into the case when the lower quality of human capital can be transient (career mobility theory) or permanent (human capital model). The analysis of the determinants of overeducation will help us discriminate among these last hypotheses. The fact that most overeducated remain such for long after entering the labor market is indirect confirmation that for most people the lack of quality of the human capital is permanent.

## 2. A survey of the empirical literature

The empirical literature on Italy has especially highlighted the low level of both demand and supply of human capital in the country. With regard to the demand side, Manacorda and Petrongolo (2000), among others, note that the production structure is still based on labour intensive traditional manufacturing. The origin of the educational mismatch could therefore be found in the weak demand for more educated workers as compared with the skill formation supplied by the educational system (Cainarca and Sgobbi, 2009). With regard to the supply side, Checchi (2003), Pastore (2012) and Franzini and Raitano (2012), among others, note the lowest level and quality of educational attainment of young people as compared to the EU average.

A large amount of literature points to the inefficiency of the educational system in generating a sufficiently high level and diversified composition of skills to satisfy labour market demand. Ordine and Rose (2009), for example, model the hypothesis that inefficient educational choices due to the different educational quality supplied by the universities can generate overeducation. It is mirrored not only in the low level of educational attainment, but also in dramatic social immobility. Tarvid's (2015) theoretical model shows that admission tests can be a better alternative to a general admission ceiling if one wants to reduce overeducation.

Educational attainment is especially low amongst the poorest segments of the population due to school tracking (see, among others, Checchi et al., 1999; Cappellari, 2004; Brunello and Checchi, 2007; Bratti et al., 2008; Checchi, 2010). In addition, Caroleo and Pastore (2012) note a strong correlation between parents' educational attainment and that of their children by field

of study, especially in those fields of study that give access to liberal professions. Despite the low level of tertiary education, the existing comparative evidence suggests a higher than average share of overeducated workers (European Commission, 2013).

Cutillo and Di Pietro (2006) find that women have a lower probability of overeducation than men in the ISTAT survey of graduates. The fields of study most at risk are political sciences, literature and languages. Instead, law, medicine, sciences, mathematics, philosophy, engineering, architecture and agriculture reduce the chance of overeducation in relation to business. Using the ISFOL-PLUS data on 2005 and 2006, Franzini and Raitano (2012) and Aina and Pastore (2012) add delayed graduation and having found their job through informal channels to the individual factors of overeducation.

Ferrante (2010) uses AlmaLaurea data to assess the impact of a number of individual characteristics on the "effectiveness of the university degree"[1] in providing a job that is in line with the educational and skill level of the individual. He reports that the variables that correlate positively and significantly with the above indicator are: a high secondary school diploma with a grade of 55 to 60 out of 60; a high university final grade; a longer length of job search; undergoing some postgraduate training; holding a university degree in engineering, chemistry, pharmacy or law. The negative and statistically significant determinants include holding a diploma from a technical high school rather than a grammar school; having parents who belong to the working class; starting their career via starter or atypical working contracts, such as apprenticeships, internships or temporary contracts; holding an arts degree or a degree in education, psychology or social sciences.

A recent stream of the literature is highlighting the different chances of employment and overeducation by field of study. by looking at the experience of graduates from the University of Padua, Boccuzzo et al. (2016) stress the importance of choosing the "right major" at the beginning of the university career. Gaeta et al. (2017) find also a strong correlation of the chances of overeducation of Italian Ph.D. holders by field of study.

As in other countries, even if the return to education is still positive for the overeducated as compared to secondary high school diploma holders (Franzini and Raitano, 2012; Cainarca and Sgobbi, 2009) the overeducated nevertheless invariably get a wage penalty compared to their peers employed in positions for which they hold the required diploma (Sloane, 2003; Leuven and Oosterbeek 2011). Moreover, generally speaking, the wage penalty for overskilling is smaller than that for overeducation (see, among others, McGuinness and Sloane, 2010).

1 This is a special indicator created by AlmaLaurea by merging answers to the two questions available in the questionnaire on the educational mismatch (A16 and A17).

The wage penalty for overeducation/overskilling is found to be lower in Italy than in other countries and, in some cases, it is found to be statistically insignificant (Wasmer et al., 2005; Ordine and Rose, 2009). Using the 2001 ISTAT survey on professional integration of graduates in 1998, Cutillo and Di Pietro (2006) find a wage penalty for university graduates ranging between 2.4% and 5.7%. Using Reflex data, McGuinness and Sloane (2010) find a wage penalty of about 10%. Interestingly, in the case of Italy, they find a higher wage penalty for the overskilled (-11%) than for the overeducated (-4%), the latter being statistically insignificant. On the contrary, using the ECHP data set, Bàrcena-Matìn et al. (2012) find that the pay penalty of the overskilled is not statistically significant in Italy. Using the ISFOL PLUS data, Aina and Pastore (2012) find a wage penalty associated with overeducation of about 20%, slightly higher than in previous studies.

After controlling for sample selection bias, most authors find that the wage penalty associated with overeducation increases with respect to the OLS baseline estimates in absolute value, which lends support to the job competition, job assignment and, above all, human capital models, as based on our interpretation of the Heckman procedure explained later. Sloane et al. (1999) and Dolton and Vignoles (2000) are, to our knowledge, among the first to use the sample selection procedure to detect the existence of omitted heterogeneity of the non-employed as compared to the employed. Interestingly, similar to us, they find only a small increase in the coefficient of the overeducation variable. Using an ISTAT survey carried out in 2001 on graduates in 1998, Cutillo and Di Pietro (2006) find only a small impact of sample selection bias on the coefficient of overeducation.

The issue of the endogeneity of overeducation due to unobserved heterogeneity of the overeducated or to other factors, such as measurement errors, is not addressed here due to the lack of appropriate longitudinal data. In the past, endogeneity has been handled in different ways in the estimates of wage penalty associated to overeducation: fixed effects regressions (Bauer, 2002; Cutillo and Di Pietro, 2006); IV estimates within the context of cross-section data (Korpi and Tåhlin, 2009); and, more recently, the direct inclusion of ability controls to catch differences in skill levels between the overeducated and the rest of the sample (Allen and Van der Velden, 2001; Korpi and Tåhlin, 2009). Kleibrink (2016) uses all of these methods.

It is common knowledge in this literature, however, that it is impossible to address appropriately endogeneity issues in the context of cross-section data. The reason why this is the case is that overeducation and wages are two sides of the same coin: a given labour market match. It Is impossible then to disentangle one from the other and there is hence no variable able

to satisfy both the conditions for the implementation of IV analysis, and especially the exclusion restriction[2].

## 3. Methodology

A simplified variety of the Verdugo and Verdugo (1989) version of the classical ORU specification (Over-, Required, and Undereducation), in which the usual Mincerian earnings equation is augmented by overeducation and/or overskilling dummies, provides the empirical framework for estimating the wage penalty[3]:

$$Lnw_i = r^{OLS} O_i + \sum_{i,j=1}^{n} \beta_j X_{i,j} + u_i \qquad [1]$$

where $Ln\ w_i$ is the natural logarithm of the net monthly wage for an individual $i$, the $X_i$ [4] are a set of control variables assumed to affect earnings and the $\beta_j$ are their coefficients. $O_i$ is a dummy equal to one when the individual $i$ is mismatched and $r^{OLS}$ is the estimated coefficient. $u_i$ is a disturbance term representing other forces which may not be explicitly measured, assumed independent of $X_i$ and $O_i$. The latter dummy is here taken according to the specification adopted to mean overeducation, overskilling or different interactions of the two or with other variables. The $X$, variables include controls for gender and marital status, nationality, type of high school diploma, the final grade at the university, the field of study, the time spent to get a degree, whether the graduate studied abroad, the type of post-graduate studies done, if any, whether they moved after graduation to find a job. The next section provides a more detailed list of regressors

Many observers have raised the concern that simple OLS estimates might tend to under/overestimate the wage penalty associated with the educational mismatch. There might be, in fact, unobserved heterogeneity between overeducated and the rest of the sample of graduates (endogeneity bias) and, at the same time, between the employed and the non-employed (sample selection bias). As Kleibrink (2016) has recently shown, the first source of bias, the endogeneity bias, is generally addressed in three ways in the overeducation literature: 1) by means of longitudinal data, which however rely exclusively on the small group of those who change their educational matching and therefore is not very reliable in the case of overeducation (Korpi and

[2] The AlmaLaurea data do have a longitudinal dimension, but it is not accessible for research purposes.

[3] Unfortunately, the AlmaLaurea data provides no information on what Duncan and Hoffman (1981) call the "surplus education" (and the "deficit education"), namely the number of years in excess (in deficit) with respect to those required for the job, which is to be preferred, according to Leuven and Oosterbeek (2011). Moreover, the data does not allow us to measure under-education.

[4] We exclude job characteristics, partly endogenous.

Tåhlin, 2009; Leuven and Oosterbeek, 2011); 2) by means of IV estimates, which, however, have also to face the lack of suitable instruments due to the strong link between the probability of being overeducated and the wage earned (Korpi and Tåhlin, 2009); 3) by controlling for the quality of human capital in the regression, as Kleibrink himself does by using different indicators of skill coming from the German Socio Economic Panel (GSOEP), and in particular the data from the International Adult Literacy Survey (IALS).

In the case of the AlmaLaurea data, unfortunately, approach 1) cannot be implemented due to the unavailability of the longitudinal dimension of the data; approach 2) has been tested in omitted estimates which are the result of considerable experimentation, using a large set of different instrumental variables none of which overcame the two conditions for being a good instrument and especially the exclusion restriction; and approach 3) is actually controlled for with a whole battery of proxies for the individual unobserved lower ability of the overeducated. The latter point is, first, addressed by studying the determinants of overeducation and by showing that overeducation is generally associated on average with lower grades at school and at the university, lower educational background in terms of school tracking and of family background, delayed graduation, lower level of graduate training and studies and so on. We expect, in fact, that when adding the available proxies for individual ability and motivation the coefficient of overeducation in the OLS estimates reduces in absolute value, therefore confirming the assumption of endogeneity bias. This is what Kleibrink (2016) finds using as regressors measures of the skills coming from the IALS: the coefficient of the overeducation variable becomes not statistically significant both in the IV and in the OLS estimates including the IALS controls for individual skills.

Several economists have proposed the Heckman (1979) sample selection procedure as an empirical model to address the issue of omitted heterogeneity of the non-employed and assess its impact on the wage effect of overeducation. Neglecting the non-employed might generate a bias on returns to education and also on the wage effect of the educational mismatch whose direction is in principle ambiguous (see, among others, Sloane et al., 1999; Dolton and Vignoles, 2000; Cutillo and Di Pietro, 2006).

In the context of estimates of the reservation wage, Nicaise (2001) suggests to interpret the sign of the bias in the estimated coefficient in Heckman sample selection corrected estimates as compared to the OLS estimates as indication in support for one of two alternative theories of unemployment, namely what he calls the “crowding hypothesis” and the “reservation wage” hypothesis based on the job search model. It is our specific contribution and the main novelty of this paper to apply his line of reasoning to the case of overeducation. We show that the

Heckman sample selection model can be used as a screening device to select some of the alternative theoretical interpretations of overeducation discussed in Section 1.

According to $TH_0$/ $TH_1$, in the job competition, the job assignment and the human capital model, sample selection bias might arise because the mismatch appears first of all in the form of a higher probability of non-employment and only at a later stage it takes the form of a wage penalty for overeducation. After controlling for the selection bias arising from considering the non-employed, hence, the wage penalty associated to overeducation should be higher. Conversely, according to the job search theoretical model, non-employment is a voluntary choice, since the most skilled graduates prefer to remain non-employed while waiting for the best job offer they can get to come. If employed, they would be less likely to experience overeducation. In this case, after controlling for the selection bias arising from considering non-employment, the wage penalty of overeducation should be lower.

OLS estimates do not control for possible unobserved differences between the mismatched and the non-employed, who might also experience the mismatch if employed. To do so, we adopt the Heckman (1979) specification – sometimes called Heckit[5] – of the earnings equation, where the usual OLS estimates are corrected for the lower/higher employment opportunities of the most skilled and motivated among those whose personal attributes would lead to overeducation if they were employed. In analytical terms, equation [1] should be specified differently:

$$Lnw_i = r^{Heckit} O_i + \sum_{i,j=1}^{n} \beta_j X_{i,j} + \rho\lambda\left(\sum_{l=1}^{m} \theta_l Z_{i,l}\right) + u_i \qquad [2]$$

where $r$ is now denoted with the superscript Heckit, to distinguish it from the corresponding OLS estimate; $\rho$ is the correlation between the error terms of the main and of the participation equation and $\lambda$ is the inverse Mills ratio evaluated at the mean of the covariates ($Z\vartheta$), which include, in addition to the $X$, also one or more instrumental variables. When there is sample selection bias, the latter term should be included in the earnings equation to obtain unbiased estimates of the parameters of interest.

Two possibilities are in order:

$$H_0: r^{Heckit} > r^{OLS} \qquad [3]$$

$$H_1: r^{Heckit} \leq r^{OLS}$$

As shown in panel (a) of Figure 1, according to $H_0$, OLS is underestimating the wage penalty associated with overeducation/overskilling. Only the most skilled overeducated (pink dots) are selected into employment. Once controlling for the least skilled and motivated among

[5] Heckit: 'Heck-' from Heckman and '-it' as in probit, tobit, and logit.

those experiencing the mismatch, the coefficient of the overeducation variable increases (green line). $H_0$ is consistent with the job competition and also the job assignment model[6], whereas unemployment is high and hence dominated by the involuntary component. The most skilled are the first to get job offers and accept them as the best alternative. We argue here that, in fact, $H_0$ is also consistent with the human capital model, in as much as selection into employment refers to the graduates with the highest quality of human capital. The least skilled would be such because of their lack of work-related competences: their lower human capital tend to relegate them into non-employment.

As shown in panel (b) of Figure 1, according to $H_1$, OLS is overestimating the wage penalty associated with the educational mismatch. Only the least skilled overeducated are selected into employment (pink dots). Once controlling for the most skilled and motivated among those experiencing educational mismatch, the coefficient of the overeducation variable shrinks (green line). $H_1$ is consistent with the search theoretical models, whereas unemployment is assumed to be voluntary in nature and the most skilled graduates prefer to wait in the non-employment pool for the best job offer to arrive.

**[Figure 1 about here]**

In addition, further empirical evidence has been also used to choose our preferred theoretical interpretation among the three selected as based on the Heckman procedure, namely the job competition, the job assignment and human capital model. Several concurrent empirical tests suggest us selecting the latter model, namely the human capital model, as the most appropriate, at least in our data, but it is likely that our conclusion apply more generally also to other countries and contexts, as previous studies also suggest similar estimated coefficients. In fact, the job competition and, to some extent, also the job assignment models, both assume that the overeducated have still the same level of human capital that the non-overeducated have. We show that this is not the case and that overeducation is associated with a lower quality of education. This is shown, above all, in the analysis of the determinants of overeducation of next section. In addition, we test whether the coefficient of overeducation is changing when including proxies for the skill level of the respondents, Kelibrink (2016) finds that after including IALS data on skills, the coefficient of overeducation becomes not statistically significant anymore, suggesting that most part of it was explained by the lower skill level of the overeducated. As Leuven and Oosterbeek (2011) also note, this would be consistent also with an interpretation in favor of endogeneity bias and, hence, of unobserved heterogeneity of the overeducated.

[6] We thank Peter Sloane for suggesting this extension to us.

The new interpretation of the Heckman procedure prompted in this paper will allow us better understanding the nature and determinants of overeducation and, hence, better defining policy instruments to fight it and prevent the waste of private and public resources that it causes.

## 4. The AlmaLaurea data

This is the first study available using AlmaLaurea data to assess the extent of overeducation. AlmaLaurea is a consortium including a large and growing number of Italian universities[7]. Gathering very detailed information on several aspects of university education and the school-to-work transition of graduates, AlmaLaurea is the most important source of information for assessing the quality of tertiary education from a comparative perspective across athenaeums, faculties, provinces, fields of studies and so on.

Our sample is made up of pre-reform graduates (so-called *"vecchio ordinamento")*, who graduated in 2005 in one of the 36 universities belonging to the consortium at that time[8]. Individuals in the sample are observed at the time of their graduation and thereafter in 2006, 2008 and 2010. Our focus is on respondents five years after graduation to study the determinants of permanent rather than transient overeducation. Our data should not be affected by the recession, the major impact of which was not felt by the Italian labour market until the end of 2011.

The sample consists of 28,976 pre-reform graduates interviewed at the time of graduation, 21,605 of whom answer the questionnaire five years after graduation and 17,387 of whom report being employed. The attrition rate of about 25% five years after graduation is relatively low for this type of longitudinal data (Table 1).

This is by far the largest and most homogeneous sample of young graduates available in the country. In addition, it is a purpose specific sample survey to assess the quality of education and, hence, it reports comprehensive and very rich information on different aspects of the production of skills. Using a data set similar to ours, Bagues and Sylos Labini (2009, Table 4.5 and 4.6) show that the sample of AlmaLaurea graduates does not differ from a statistical point of view from the universe of university graduates, which, in turn, suggests that the AlmaLaurea data is representative of the underlying population under many different individual level

[7] As at 2016, 72 Italian universities were members, representing 91% of all Italian graduates. For further details about AlmaLaurea, see its homepage: http://www.almalaurea.it.

[8] We do not include in the sample students enrolled after the implementation of the 3+2 reform in 2001 (the so-called Bologna reform), because only few of them, the most skilled, have graduated. In fact, due to the widespread phenomenon of delayed graduation, very few graduates would be included in the sample.

(gender, age, high school grade) and institutional differences, as measured by the number of students per university and per professor, as well as by the share of delayed students. The Authors provide detailed description of the aims and way of functioning of the AlmaLaurea data bank, which has become, over time, an important term of reference for many other similar consortia in Europe and elsewhere.

Most previous studies have used the 2001 ISTAT survey of graduates which is based on a much smaller sample (about 14,383 graduates) in 1998 and has a much less detailed questionnaire. Most other Italian data sets previously used to study overeducation are general sample surveys covering the entire active labour force, with a very heterogeneous sample of individuals of all ages, in different stages of their working career, and, on the top of that, having a much smaller sample size. For instance, the ISFOL Plus data bank generally includes a sample of about 16,000 individuals, of which about 3,000 are graduates over the entire population.

The employment questionnaire conducted after graduation includes two questions that provide subjective measures of the educational mismatch. Question A16 asks: "In your current job, do you use the competences acquired during your university studies?" Three answers are possible: 1) the competences acquired are used to a great extent; 2) they are little used; 3) they are not used at all. We defined those who choose answer 3 as overskilled. This question closely mirrors what Dolton and Silles (2008) call the "to do" definition of the mismatch. Question A17 asks: "Is your university degree necessary to obtain your current job?" Four answers are possible: 1) the degree is required by law; 2) it is not required by law, but is in fact needed; 3) it is not required by law, but is in fact useful; 4) it is neither required by law nor useful. We defined all those who choose answer 4 as overeducated. This question allows us to define what Dolton and Silles call the "to get" definition of the mismatch.

Earnings are defined as the natural logarithm of net monthly wages. Question A20 asks the interviewee to indicate which of 13 classes of €250 of monthly earnings up to the "over €3,000" class, (s)he belongs to. For ease of analysis, the natural logarithm is applied to the average value of the relative class. In our earnings estimates, we use both the OLS and interval regression method. No information on working hours is available in the data we use.

Questions about such individual characteristics as civil status are only asked at the time of graduation, not five years later. This ensures the exogeneity of these variables, but cannot prevent them from being inaccurate.

The independent variables are self-explanatory. They have been grouped in: a) individual characteristics (gender, civil status, having children, nationality); b) educational background (type of high school); c) university attendance and performance (final grade, time of graduation, field of study); d) pre-graduation work experience (experience of study abroad, work and

study); e) post-graduate studies and training; f) whether the graduate moved from the place where they obtained their degree to find their job.

The AlmaLaurea data has a number of advantages as compared to the two main concurrent data sets used in Italy (ISTAT and ISFOL Plus):

a) overall, the sample included about 20% of the pre-reform graduates in 2005, the largest possible number for a homogeneous cohort;

b) it covers a more recent period, but just before the current economic crisis began;

c) it is obtained by merging two extremely comprehensive datasets, of which the first is elicited at the time of graduation and contains all types of information on the study career, both in secondary and tertiary education, and the second contains extensive information on post-graduate training and early labour market experiences;

d) it allows the definition not only of overeducation, as in previous studies concerning Italy, but also overskilling.

# 5. Results

## 5.1. Size and composition of the sample

Table 1 shows that one year after graduation, the overskilled and the overeducated amounted to about 16.5% and 13.2% respectively, falling at a roughly constant rate to 11.4% and about 8.0% respectively at the end of the period under consideration. It means a reduction down to only about 69% and 61% of the original value[9]. In other words, the educational mismatch is *not* a transitory phenomenon for a large number of individuals, which is in line with the findings of a growing body of literature (Mavromaras et al., 2015; Congregado et al., 2016; Baert et al., 2013; Baert and Verhaest, 2014).

**[Table 1 about here]**

The descriptive evidence highlights a massive and quite generalized disruption of the human capital that the university system has generated. Figure 2 reports the non-employment

[9] Following the suggestion of one of the anonymous referees, we tried to assess whether this reduction in the share of overeducated and overskilled from 1 to 5 years after graduation was due to attrition and namely a higher than average rate of dropout from the survey of the overeducated. We have found that the share of dropouts from the survey by labor market status between 1 and 3 years, between 3 and 5 years and between 1 and 5 years is distributed quite similarly across labor market statuses, with only a slightly higher probability to dropout for the individuals who are both overeducated and overskilled. This suggests that the reduction in the share of overeducated / overskilled is not due to attrition.

shares by field of study, the share of overeducation and that of the well-matched graduates five years after obtaining a degree. The non-employment share is extremely high for most degrees, with an average of 40%. It ranges between a minimum of 27% in the case of engineering and a maximum of about 50% in the case of geology and biology. The low employment share of graduates in medicine is due to the fact that most medical doctors are still attending postgraduate schools.

A more accurate measure of the human capital waste is given by the sum of non-employment and overeducation/overskilling (Baert and Verhaest, 2014). Interestingly, it is quite common that the higher the non-employment share is, the higher is the share of the overeducated.

Figure 2 displays the share of employed graduates who are overeducated and overskilled for each field of study, which appears to be undoubtedly one of the main factors, as also other studies have highlighted on Italy (Cutillo and Di Pietro, 2006; Boccuzzo et al., 2016; Gaeta et al. 2017) and elsewhere (Cappelli, 2015; Congregado et al. 2016). Overeducation ranges from zero or almost zero in the case of medicine, architecture, chemistry and pharmacy, engineering and sciences to more than 10% in the case of geology and biology (10.2%), physical education (12.2%), languages (13.2%), political and social sciences (14%) and literature (17.9%). Overskilling follows roughly the same pattern with a share slightly higher for each field of study. Languages (16.5%), political sciences (18.4%), geology and biology (18.7%), physical education (20.7%) and literature (25%) present the highest percentages.

When looking at the productivity characteristics of the overeducated and overskilled versus the perfectly matched[10], the former appear to have much lower quality of human capital, both in terms of education and work-related skills, which we explore in more depth in the next section.

**[Figure 2 and 3 about here]**

## 5.2. Determinants

Table 2 reports odds ratios of the independent variables estimated by logit on the probability of being overeducated (column 1) or overskilled (column 2). Ceteris paribus, gender is a statistically significant (at a level of 5%) determinant of overskilling, but not of overeducation. Women are about 13 odds points more likely to be overskilled. Unreported estimates confirm that women have a statistically significant and higher non-conditional probability of both overeducation and overskilling, by about 15 and 36 odds points,

[10] For the sake of brevity, we omit the table which is available on request from the authors.

respectively. The gap dramatically shrinks when we include controls for human capital variables, pointing to their tendency to concentrate in those fields of study which are associated with a greater mismatch. Remarkable and stable gender segregation by fields of study is a well-known characteristic of Italy and other countries (see, for instance, Triventi, 2010).

Other individual characteristics, such as the civil status and having children at the time of graduation, seem to have little impact on the probability of being overeducated, probably because they might have changed five years from graduation.

The statistical significance of high school performance seems to confirm what a large amount of literature (see, among others, Brunello and Checchi, 2007; Caroleo and Pastore, 2012) says about the role of the family socio-economic background in affecting, through school tracking, the labour market performance of graduates. The same groups that are at a disadvantage in achieving higher education are also at a disadvantage in their access to the labour market and, we add here, frequently tend to experience educational mismatch. In fact, despite being completely free, the choice of the type of high school tends to reflect the social class of origin. People from a poor walk of life tend to choose technical or professional schools and, hence, experience problems in their educational career later on, also augmenting the chance of being overeducated in their working life. For these reasons, after controlling for their performance at high school, it should come as no surprise that the educational background of parents does not directly affect the probability of being mismatched. We have therefore excluded these variables in our final estimate.

Several aspects of an individual's educational quality correlate with the likelihood of experiencing a mismatch (as noted in Leuven and Oosterbek, 2011). This is the case with the aforementioned field of study, the final grade and the time spent obtaining a degree. *Fuoricorso* graduates with a delay of two or more years have ceteris paribus a 50-odds-point greater chance than the *in corso* of experiencing educational mismatch (on this also see Aina and Pastore, 2012).

The impact of the field of study is particularly important. All fields of study are associated with a higher chance of mismatch than engineering (the reference group). Particularly strong is the impact of holding a degree in literature, languages, physical education, political and social sciences, psychology and geology and biology. Only architecture and medicine are not statistically different from engineering. Overall, the quality of education, as measured by indicators of university performance, seems to be the most important determinant in the probability of being mismatched.

The localization of job search matters. The graduates who seek their job in the north, no matter whether west or east, experience a much lower probability of mismatch than their peers

in the centre and even more so in the south. Moving abroad reduces the risk of overeducation, but not of overskilling, probably because of the country-specific content of some of the skills acquired in the educational system, which makes them less easy to transfer.

Interestingly, on-the-job training practices, attending some graduate schools and master degrees of level II reduce the risk of mismatch in a statistically significant manner. Masters of level I reduce the risk of overeducation, but not overskilling, which are positively affected by post-degree scholarships. Other post-degree programmes – such as the doctorate, other types of master degrees, internships, public training programmes and voluntary social work – are not statistically significant.

Overall, our findings seem to indirectly confirm rather than contradict the human capital model. In fact, the mismatch is strongly associated with characteristics denoting a low quality of human capital. In other words, while increasing in the educational dimension, the human capital of graduates does not sufficiently develop in terms of the skills which are directly requested in jobs for graduates. In other words, there is a demand for job-specific skills in the labour market which remains unsatisfied.

**[Table 2 about here]**

As a robustness check we have also estimated the Heckprobit model of the determinants of overeducation / overskilling, since many individuals observed in 2010 are non-employed and this might possibly bias the coefficients of key regressors. The Heckprobit model is similar to the Heckit model explained above, but it has a probit model not only as selection equation, but also as main equation (Van de Ven abd Van Pragg, 1981). This is necessary when the dependent variable is a dummy, like overeducation / overskilling. In omitted estimates of the Heckprobit model[11], we do find evidence of a statistically significant correlation between the main (probit of overeducation / overskilling) and selection equation (probit of employment / non-employment), since the estimated athrho is statistically significant in both estimates, relative to overeducation and overskilling, but the coefficients of key variables have the same statistical significance of the ones presented in the logistic regression (Table 2), while being not easy to compare with the simple probit model in terms of absolute values, due to the high non-linearity of the probit model. Due to the irrelevance of the Heckprobit model in this case, we find it more meaningful to leave in the final version the logit estimates, which have the advantage of being of easier interpretation once transformed in relative risk ratios. In fact, the role of these estimates here is to show essentially the strong correlation between the probability of being overeducated / overskilled and the fact of holding a low quality of human capital. This point is more easily made using the simple logit model.

[11] The estimates are available on request.

## 5.3. The wage effect

Table 3 provides summary measures of the wage gap, in terms of the average wages, as well as in terms of unconditional and conditional estimates, by OLS and interval regression[12], and different types of estimates corrected for sample selection bias by using different forms of the Heckman procedure (as based on Table 4)[13]. The unconditional wage penalty is relatively high for both overeducation and overskilling, considering the very low returns to education in Italy. OLS slightly underestimates the wage penalty as compared to interval regression, which is the most appropriate one considering that wages are not defined in the continuum, but in intervals. However, after controlling for the level and quality of human capital, both OLS coefficients are about halved, suggesting that most part of the unconditional wage penalty is, in fact, related to measures of skill characteristics that the overeducated/overskilled possess less than the rest of the sample. Similar reductions are observed in the case of interval regressions. We report different conditional estimates, of which the former do not include some of the measures available of the quality of human capital, which are instead included in the latter[14]. We refer in particular to: the final grade reported at the university; whether they delayed their graduation beyond the typical curricular years; whether they had some study abroad experience; whether they had some form of off- or on-the-job training; whether they attended a graduate studies' program. The coefficients measuring the wage effect of overeducation and also overskilling are only marginally affected by the inclusion of these variables, though. This suggests that such measures of the quality of education cannot be compared to those available, for instance, in the GSOEP, so called IALS, used, for instance, by Kleibrink (2016). The Author himself did not find any effect from including in the estimates more general measures of the quality of education coming from the GSOEP data.

**[Table 3 about here]**

## 5.4. The Heckit estimates

Table 4 reports the results of the Heckit earnings equations specification as obtained by maximum likelihood simultaneous estimates, rather than the two-step procedure. It is the result

[12] OLS and interval regressions are available on request from the authors.

[13] The earnings equations include only 16,591 graduates because 796 workers do not report their wage, either out of apathy or because it is too low to report.

[14] We thank one of the anonymous referees of the Journal for suggesting this further robustness check.

of significant experimentation with different types of instruments and specifications. Results relative to different combinations of the mismatch are given in the last columns of Table 4.

We apply the rule by which variables in the main and selection equation should be the same, except for some instrumental variables (Cameron and Trivedi, 2009). The latter should affect the probability of participating in the labour market, but not wages (exclusion restriction). The typical IVs used in these cases, namely civil status and having children, influence wages as well, although they are not statistically significant determinants of the mismatch (Table 2). This prevents us from using these as IVs. We include them, instead, in both equations.

As an alternative, we use the educational level of parents, based on the assumption that differences among graduates in the probability of finding a job are essentially linked to the socio-economic background of parents. This is a consequence of the weakness of other tools of job search that are able to equalize chances in the labour market[15].

The arthrho variable, which indicates the correlation between the two equations, is negative and statistically significant, although to a lesser extent in the case of overskilling. The Wald test of independence between the main and selection equation confirms this result. In fact, it rejects the hypothesis $H_0$ with a high level of significance in the case of overeducation, but only at a 10% significance level in the case of overskilling. Overall, there are unobserved factors that affect the labour participation chances and, consequently, the reservation wage and also the wage received if employed. The same omitted result is obtained with the two-step estimates. Following Ermisch and Wright (1994), we do not attribute a special economic meaning to this negative coefficient of arthrho, since it can be due to several statistical properties of the estimated Heckman equation.

In the absence of other plausible instruments, we have also exploited as an instrument the strong non-linearity of the ML function, as Cameron and Trivedi (2009) suggest as an alternative to using instrumental variables which do not fully respect the two requirements of exogeneity and may, hence, generate bias of an unknown nature (Table 3 and 4). Unreported results again suggest that there is sample selection bias in the same direction of an increase in the coefficients.

**[Table 4 about here]**

However, similarly to previous studies (Sloane et al., 1999; Dolton and Vignoles, 2000; Cutillo and Di Pietro, 2006), the wage penalty associated with overeducation/overskilling increases only by about 1% in both cases. This finding could be partly due to the inadequacy of the available instruments in fully correcting for sample selection bias or also to the extreme flatness of the (entry) wage distribution, especially among a homogeneous sample like the one

[15] In omitted estimates, available on request, the educational background of parents does not affect wages.

considered here. Taken at face value, though, our findings overall can be seen as evidence in support of the job competition and job assignment models instead of the job search model. In fact, it is also in line with our interpretation of Italian overeducation as based on the human capital model. In other words, the non-employed might have a low human capital level, supposedly, the work-related component. This conclusion is in line with the analysis of Section 5.2 regarding the determinants of overeducation, which points to the low quality of human capital of the overeducated.

## 5.5. Extensions and robustness checks

This section aims to extend previous findings and, at the same time, run some robustness checks. First, we test whether the impact of the educational mismatch n earnings is due more to overeducation only, overskilling only or to both, along the lines laid down in Mavromaras et al. (2013) and Pecoraro (2014)[16]. The case when an individual is, at the same time, overeducated and overskilled is called, sometimes, genuine overeducation, since only when the overeducated is also overskilled she is really experiencing a mismatch. First of all, we provide some summary statistics. Similar to Figure 2, Figure 4 reports the breakdown of the individuals experiencing one of these three type of mismatch by field of study. It shows that genuine overeducation is particularly sizeable for graduates in the fields where also overeducation has been found to be more sizeable, namely Physical education, Geology and Biology, Arts, Languages, Political and Social Sciences, and Psychology[17].

**[Figure 4 about here]**

Table 5 reports coefficients of the wage penalty for overeducation only, overskilling only and overeducation and overskilling as based on different specifications of the earnings equation. Confirming the finding of the previous literature, the estimates suggest that those experiencing both forms of mismatch also report the greatest wage penalty, both the unconditional (23.9% and 29.1%) and the conditional one: 13% with OLS and 16% with IR. Overeducation only also bears a relatively high wage penalty. Overskilling only, instead, is associated to a small wage penalty. In fact, the coefficient is not statistically significant in OLS estimates, while it still bears a 5% wage penalty in interval regressions. In all cases, the Heckman correction implies an increase in the coefficient which does not allow rejecting the $H_0$ of job competition, job assignment and human capital model.

[16] We thank Peter Sloane for suggesting us this extension.

[17] Since more individuals experience at least one form or another of the mismatch, the other components (non-employment and well-matched) are slightly smaller than in Figure 2.

**[Table 5 about here]**

Next step consists of looking at the robustness of the restrictive definition adopted for overeducation and overskilling. Such definitions provided in the AlmaLaurea data are based on not perfectly comparable questions: the questions A16, for overskilling, and A17, for overeducation[18]. In particular, on the one hand, it is understandable the approach of defining overeducated those individuals answering to the Question A17 "it is neither required by law nor useful", and not overeducated those answering to the Question A17 "it is not required by law, but is in fact useful", On the other hand, though, classifying as not overskilled individuals answering to the Question A16 "they are little used", is quite different from the corresponding item in A17 "but is in fact useful". To test the stability of estimation results, it could be worth defining overskilled those individuals answering to the question A16 both item 2) and 3). We start from providing in Figure 5 the distribution of answers A16 (Panel (a)) and A17 (Panel (b)): it is apparent that the share of overeducation goes up dramatically in both cases, namely up to about 40% for overeducation when one considers the answers 3) and 4) to question A17: and up to about 50% for overskilling when we consider together the answers 2) and 3).

We report in Model (2) in Table 5 also the results of estimates relative to the new definition of overskilling. As to be expected, comparison of the findings of Model (2) in Table 5 and of Column (2) in Table 3 shows that the extension of the group of the overskilled has further reduced the wage penalty associated to overskilling. This is quite obvious as it mirrors the tendency of the overskilled to be confused with the rest of the graduates. It was also one of the reasons why we eventually opted for the more restrictive definition. In fact, also the size itself of the phenomenon would have been to big to be true, although the papers who adopt the statistical or objective definition of overeducation return a higher share than that obtained applying the subjective definition (for a recent comparison of outcomes of these two definitions, using the ISFOL PLUS data, see Mandrone et al. 2016).

A last robustness check consists of introducing in the econometric specifications adopted for the earnings equation a demand-side control variable, possibly accounting the different distribution of economic activities by local area. We have chosen to control for the local unemployment rate at the time when the data was collected with a procincial level of aggregation (NUTS3 level). We use the provincial unemployment rate of the area where university graduates search for a job in OLS and IRs, but the provincial unemployment rate of the area of residence for the Heckit estimates, since the destination area of non-employed individuals is not defined. To take into account the Moulton (1986) problem we compute standard errors by clustering observations at a provincial level. The unreported coefficient of the

[18] We thank one of the anonymus referees for suggesting us this robustness check.

provincial unemployment rate in the OLS earnings equations equals -0.0228 and is statistically significant at the 1% significance level. It is more than double the usual coefficient of the wage curve (-0.01), suggesting that the graduate labor market is much more flexible to changes in the local unemployment rate. However, the wage penalty associated to overeducation and overskilling changes only marginally when controlling for the local unemployment rate.

## 6. Discussion, concluding remarks and policy implications

This paper has studied the main determinants and labour market effects of the educational mismatch in Italy using, for the first time, the AlmaLaurea database. This is by far the largest and most detailed database concerning university graduates in Italy. The sample includes the universe of pre-reform graduates, namely graduates enrolled before the Bologna reform was implemented, from the universities belonging to the consortium in 2005. The focus is on labour market outcomes five years from graduation. The data allows us to establish a number of links with overeducation that have never been investigated before with the same detail. Results show that overeducation and overskilling are persistent phenomena also long after graduation for a large majority of graduates.

The evidence provided in this paper is consistent with a novel interpretation of the educational mismatch in Italy. The usual interpretation is that it is due to a lack of demand for human capital because the country is still heavily specialized in the traditional manufacturing industry, but is experiencing a dramatic increase in the percentage of graduates (Ferrante, 2010; Franzini and Raitano, 2012). This view is questioned in our analysis on the grounds of the inefficiencies of the tertiary educational and training system, particularly in enhancing job-related skills in graduates, as also evidenced in Cipollone and Cutillo (2013).

Our interpretation is consistent with the recent theories of the educational mismatch that trace it back to overly low human capital, rather than an excess of it, since, in spite of the growing level of youth general education, the work-related skills and competences continue to be insufficient (Leuven and Oosterbeek, 2011). We do not overlook the importance of the demand side. There is a potential demand for skills in the production system which remains unexploited because of the youth experience gap and their educational mismatch (Pastore, 2012; 2014).

Our conclusion is based, above all, on analysis of the determinants of overeducation. The factors associated with the mismatch are consistent with the well-known image of an immobile social structure, whereas not only success at school and at university, but also in the labour market dramatically depends on the socio-educational background of young people, which

strongly affects the choice of the field of study (Checchi et al., 1999; Caroleo and Pastore, 2012). The chances of overeducation/overskilling are strongly associated with any other university degree than engineering, medicine and a few others. Particularly strong is the impact of holding a degree in social sciences, but also in some scientific fields, such as geology and biology.

On the other hand, having completed some post-graduate training or advanced master courses, especially those involving on-the-job training, represents a cushion against the risk of overeducation, confirming the lack of job-specific skills of graduates.

The correlation between overeducation, but even more so overskilling, on the one hand, and weak educational background and/or poor university performance, on the other, in fact suggests that overskilling is an indication of low not high skills. In other words, even when there is the right match between the qualification held by the graduate and that required to get the job, perhaps due to scant work experience, some graduates might be given tasks for which they feel overskilled.

The unconditional wage penalty associated with the educational mismatch is higher than that found in other similar studies, relatively more for overeducation (between -21% and -25%) than for overskilling (between -16% and -21%). Correcting for the observable characteristics available in the AlmaLaurea database in a multivariate context, the wage gap associated with both forms of educational mismatch halved. This mirrors the low human capital endowment of the mismatched as compared to the well-matched, which also explains the wage gap to a large extent.

The Heckit correction has been used to control for the possible sample selection bias arising from measuring overeducation only among the employed without considering the different characteristics of the non-employed, which modify the probability of being overeducated. The Heckit correction confirms that there is positive selection into employment of the most skilled among those experiencing the educational mismatch, as in the job competition and job assignment models. We argue that this conclusion is also consistent with the human capital model. Nonetheless, in our data, the wage penalty only increases by about 1%. This may be due to the high youth unemployment rate, which weakens the selection mechanism or also to the flatness of the entry wage distribution.

Overall, the findings of this paper have important policy implications. From the demand side, they suggest that the most important strategy for reducing the share of overeducation and the wage penalty associated with it would be for the country to move away from a low towards a high road to development. Overall, this is the most important strategy for accommodating the

increasing supply of human capital from the youngest generation and reducing the share of graduates who possess attributes that are not in demand in the labour market.

In relation to this point, it is important to mention the small average size of Italian firms. Small firms do not manage human resources in such a way as to fully harness and develop them. This is due to their typically informal production structure, the low propensity to delegate functions to managers and the lack of on-the-job training programmes. Schivardi and Torrini (2010) note a negative correlation between the human capital of firms' owners and managers, on the one hand, and the share of graduates in the firm, on the other.

From the supply side, it is important to: a) increase the quality of tertiary education and of human capital in general; b) reduce the length of studies for the individuals coming from poor family backgrounds in order to reduce the impact of school tracking on university success; c) provide more guidance for families and students when deciding upon their field of study at university; d) also provide vocational education and training at university level, especially for individuals from poor family backgrounds (German solution). This implies adopting the dual principle on a large scale and providing on-the-job training before or soon after the university degree; e) fully implement the Bologna process.

Future research will take advantage of the longitudinal structure of the data, when available, to account for sample selection bias in this type of data. In this direction, it could be fruitful to apply a simple test for selection bias in either random or fixed effects models, and also present a correction method based on Chamberlain's approach to panel data models, as proposed in Wooldridge (2010, p. 583).

**Compliance with Ethical Standards:**

**The authors declare that they have no conflict of interest.**

## References


Aina C and Pastore F (2012) Delayed Graduation and Overeducation: A Test of the Human Capital Model versus the Screening Hypothesis. *IZA discussion paper* 6413. March.

AlmaLaurea *XII Rapporto sulla condizione occupazionale dei laureati. Investimenti in capitale umano nel future di Italia ed Europa*. Bologna: Il Mulino pp 101-115.

Albrecht, J., Vroman, S., 2002. A matching model with endogenous skill requirements. International Economic Review 43, 283–305.

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., Verhaest, D. (2013): Overeducation at the start of the career: Stepping stone or trap? Labour Economics, 25, 123-140.

Baert, S., Verhaest, D. (2014): Unemployment or Overeducation: Which is a Worse Signal to Employers? IZA Discussion Papers, 8312.

Bagues, Manuel F. and Mauro Sylos Labini (2009), "Do Online Labor Market Intermediaries Matter? The Impact of AlmaLaurea on the University-to-Work Transition," NBER Chapters, in: Studies of Labor Market Intermediation, p. 127-154.

Barcena-Martín E, Budría S and Moro-Egido AI (2012) Skill mismatches and wages among European university graduates. *Applied Economics Letters* 19(15): 1471-1475.

Bauer, T. K. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review*, 21(3): 221-229.

Becker, G. (1962). Investment in human capital. A theoretical analysis. Journal of Political Economy, 70(5): 9–49.

Boccuzzo, G., L. Fabbris and O. Paccagnella, (2016), "Job-major match and job satisfaction in Italy", *International Journal of Manpower*, 37(1): 135–156.Bratti M, Checchi D and De Blasio G (2008) Does the Expansion of Higher Education Increase the Equality of Educational Opportunities? Evidence from Italy. *Labour* 22 (special issue): 53-88.

Brunello G and Checchi D (2007) Does School Tracking Affect Equality of Opportunity? New International Evidence. *Economic Policy* 22(52): 781-861.

Büchel F, De Grip A and Mertens A (eds.) (2003) *Overeducation in Europe*. Cheltenham. Edward Elgar.

Büchel, Felix and Mertens, Antje(2004) 'Overeducation, undereducation, and the theory of career mobility', Applied Economics, 36(8): 803 — 816

Cainarca GC and Sgobbi F (2009) Quanto paga studiare: la relazione fra istruzione e retribuzione in Italia. *Economia Politica*, XXVI(1): 53-84.

Cameron AC and Trivedi PK (2009) *Microeconometrics Using Stata*. College Station TX. , New York: Stata Press.

Cappelli, P. H. (2015): Skill Gaps, Skill Shortages, and Skill Mismatches: Evidence and Arguments for the United States. ILR Review, 68, 251-290.

Caroleo FE and Pastore F (2012) Talking about the Pigou Paradox. Socio-Educational Background and Educational Outcomes of AlmaLaurea. *International Journal of Manpower* 33(1): 27-50.

Carroll, D., & Tani, M. (2013). Over-education of recent higher education graduates: new Australian panel evidence. *Economic of Education Review, 32,* 207-218.

Cappellari L (2004) The Effects of High School Choices on Academic Performance and Early Labour Market Outcomes. *Royal Economic Society Annual Conference* 92. Royal Economic Society.

Checchi D (2003) The Italian Educational System: Family Background and Social Stratification. In ISAE (ed.) *Monitoring Italy*. Rome: ISAE.

Checchi D (ed.) (2010) *Immobilità diffusa: Perché la mobilità intergenerazionale è così bassa in Italia*. Bologna: Il Mulino.

Checchi D, Ichino A and Rustichini A (1999) More Equal but Less Mobile? Education Financing and Intergenerational Mobility in Italy and in the US. *Journal of Public Economics* 74(3): 351-393.

Cipollone Pand Cutillo A. (2013) Overeducation or asymmetric information?. *Rivista Italiana di Economia Demografia e Statistica*, LXII 2: 123-138.

Congregado, E., Iglesias, J., Maria Millan, J. (2016), "Incidence, effects, dynamics and routes out of overqualification in Europe: a comprehensive analysis distinguishing by employment status", Applied Economics, 48(5), 411-445.

Cutillo A and Di Pietro G (2006) The effects of overeducation on wages in Italy: a bivariate selectivity approach. *International Journal of Manpower* 27(2): 143-168.

Dolado, J., Jansen, M., Jimeno, J., 2009. On-the-job search in a matching model with heterogeneous jobs and workers. The Economic Journal 119, 200–228.

Dolton P and Silles MA (2008) The Effects of Overeducation on Earnings in the Graduate Labour Market. *Economics of Education Review* 27(2): 125-139.

Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the UK graduate labour market. Economics of Education Review, 19(2): 179-198.

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

Ermisch, J.F. and R.E. Wright (1994), "Interpretation of negative sample selection effects in wage offer equations", *Applied Economics Letters*, 1(11): 187-189.

European Commission (2013), *Employment and Social Development in Europe 2012*, EU, Brussels.

Ferrante F (2010) «Educational / skill mismatch»: cosa ci dicono i dati AlmaLaurea?. In AlmaLaurea (op. cit.).

Franzini M and Raitano M (2012) Few and underutilized? Overeducation of Italian graduates. In Mandrone E (ed.) *Labour Economics: PLUS Empirical Studies*. ISFOL, Temi e Ricerche 3. Cava dè Tirreni: Ediguida.

Gaeta, G.L., G. Lubrano Lavadera and F. Pastore (2017), "Much Ado Much ado about nothing? The wage penalty of holding a Ph.D. degree but not a Ph.D. job position", forthcoming in *Research in Labor Economics*.

Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica* 47(1), 153-62.

Kleibrink, J. (2016): Inept or Badly Matched? Effects of Educational Mismatch in the Labor Market. Labour, 30(1): 88-108.

Korpi T. and Tåhlin M. (2009) 'Educational Mismatch, Wages, and Wage Growth: Overeducation in Sweden, 1974–2000', *Labour Economics* 16(2): 183–193.

Leuven E and Oosterbeek H (2011) Overeducation and Mismatch in the Labour Market. In Hanushek EA, Machin S and Woessmann L (eds) *Handbook of the Economics of Education* (vol. 4). London: Elsevier, pp. 283-326.

Manacorda M and Petrongolo B (2000) Skill Mismatch and Unemployment in OECD Countries. *Wirtschaftspolitische Blatter*, 47(1): 72-82.

Mandrone, E., F. Pastore and D. Radicchia (2016), "Overeducation after the Supply Push in Italy. A Fresh Look at Five Indicators and a Proposal of a New Opposite Indicator", in A. Ricci (ed.), *Crisi economica, lavoro e imprese. Il ruolo del capitale umano in Italia*, ISFOL, Rome.

Mavromaras, K. G., McGuinness, S., O'Leary, N. C., Sloane, P. J., Wei Z. (2013). Job mismatches and labour market outcomes: Panel evidence on university graduates, Economic Record, 89(286): 382-395.

Mavromaras, K., Sloane, P., Wei, Z. (2015): The scarring effects of unemployment, low pay and skills under-utilization in Australia compared, Applied Economics, 47, 2413-2429.

McGuinness, S. (2006), "Overeducation in the labour market", Journal of Economic Surveys, 20(3): 387-418.

McGuinness S and Sloane PJ (2010) «Educational / skill mismatch» nel mercato del lavoro dei laureate: un'analisi comparata. In AlmaLaurea (*op. cit.*).

Moulton, B. (1986), "Random Group Effects and the Precision of Regression Estimates," Journal of Econometrics, 32(3): 385–397.

Nicaise I (2001) Human capital, reservation wages and job competition: Heckman's lambda reinterpreted. *Applied Economics* 33(3): 309-15.

OECD (various years), *Education at a Glance: OECD indicators. Country note: Italy*. Paris: OECD Publishing,.

Ordine P and Rose G (2009), "Overeducation and Instructional Quality: a Theoretical Model and Some Facts", *Journal of Human Capital*, 3(1): 73-105.

Pastore F (2012) Le difficili transizioni scuola lavoro in Italia. Una chiave di interpretazione. *Economia dei servizi* 7(1): 109-128.

Pastore F (2014), *The Youth Experience Gap. Explaining National Differences in the School-to-Work Transition*, Physica Verlag, Heidelberg

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*, 28(3): 309-337. Sattinger, M. (1993) 'Assignment Models of the Distribution of Earnings', Journal of Economic Literature 31(2): 851–80.

Schivardi F and Torrini R (2010), "Structural Change and Human Capital in the Italian Productive System", Giornale degli economisti, 69(3): 119-167.

Sicherman, N. and O. Galor (1990), "A Theory of Career Mobility", *Journal of Political Economy*, 98(1): 169-192.

Sloane P (2003) Much ado About Nothing? What does the Overeducation Literature really Tells us?. In Büchel F, et al. (op. cit.).

Sloane, P. J., Battu, H., & Seaman, P. T. (1999). Overeducation, undereducation and the British labour market. Applied Economics, 31(11), 1437-1453.

Tarvid, A. (2015): The effectiveness of access restriction to higher education in decreasing overeducation. Economic Analysis and Policy, 45, 11-26.

Thurow, L. (1979) ‘A Job-Competition Model’, pp. 17–32 in M. Piore (ed.) Unemployment and Inflation. White Plains, NY: ME Sharpe.

Triventi M (2010) Something Changes, Something Not. Long-Term Trends in Gender Segregation of Fields of Study in Italy, *Italian Journal of Sociology and Education* 2(2): 47-80.

*Van de Ven, W.P.M.M., and B.M.S. Van Pragg. 1981. “The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection.” Journal of Econometrics 17(2): 229–252.*

Verdugo RR and Verdugo NT (1989) The Impact of Surplus Schooling on Earnings: Some Additional Findings. *The 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.

Wasmer E, Fredriksson P, Lamo A, Messina J and Peri G (2005), *The Macroeconomics of Education*. Report for the Fondazione Rodolfo De Benedetti. Milan:FDB.

Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge MA.

# Appendix of Tables and Figures

**Table 1. Overeducation after 1, 3, 5 years of pre-reform graduates**

| Definition | 1 year | 3 years | 5 years |
|---|---|---|---|
| **Overskilled ("to do" definition)** | 16.47 | 12.49 | 11.44 |
| **Overeducated ("to get" definition)** | 13.16 | 9.37 | 7.99 |
| **Number of wage employees** | 13,500 | 17,223 | 17,387 |
| **Number of interviewees** | 25,196 | 23,851 | 21,605 |

*Source: own elaboration of AlmaLaurea data.*

**Table 2. Determinants of overeducation 5 years after graduation. Pure pre-reform graduates in 2005. Log odds ratios from logit estimates**

| Dependent variable | Overeducation ("to get") | Overskilling ("to do") |
|---|---|---|
| Independent variables | (1) | (2) |
| Individual Characteristics | | |
| Gender. Default: Men | | |
| Woman | 0.9112 | 1.1277** |
| Civil Status, Male. Default: Single | | |
| Married | 1.3706 | 1.1783 |
| Partner | 1.1641 | 0.9821 |
| Separated, divorced, widow | 0.564 | 0.9668 |
| Civil Status, Female. Default: Single | | |
| Married | 0.9802 | 0.9523 |
| Partner | 1.0728 | 0.92 |
| Separated, divorced, widow | 0.727 | 0.3719** |
| Number of children, Men | 0.5841* | 0.7369 |
| Number of children, Women | 0.8543 | 1.051 |
| Non Italian | 1.0766 | 1.0438 |
| High School | | |
| Type of high school diploma. Default: Science-based grammar school in | | |
| Classical grammar school | 0.9079 | 1.0213 |
| Specialisation in teacher training | 0.7900* | 1.003 |
| Language high school | 1.2582* | 1.0705 |
| Art school | 1.7490*** | 1.3975* |
| Technical school | 1.3541*** | 1.1776** |
| Professional school | 1.5591*** | 1.6706*** |
| Other high school diploma | 1.4884* | 1.2664 |
| University performance | | |
| Final grade at university. Default: Magna cum laude | | |
| 66-90 out of 110 | 1.9540*** | 2.2542*** |
| 91-100 out of 110 | 1.8596*** | 1.9335*** |
| 101-105 out of 110 | 1.5401*** | 1.8858*** |
| 106-110 out of 110 | 1.4571*** | 1.5971*** |
| Time spent obtaining a degree. Default: Curricular years | | |
| 1 extra-curricular year late | 1.3314* | 1.3141* |
| 2 extra-curricular years late | 1.5746*** | 1.3815** |
| 3 extra-curricular years late | 1.5557** | 1.6536*** |
| 4 extra-curricular years late | 1.5520** | 1.5552*** |
| 5 extra-curricular years late | 1.9378*** | 1.8258*** |
| Field of study. Default: Engineering | | |
| Agriculture | 4.4684*** | 3.7624*** |
| Architecture | 1.339 | 1.3471 |
| Economics and statistics | 3.4690*** | 1.6251*** |
| Physical education | 9.0089*** | 7.8656*** |
| Geology and biology | 7.0864*** | 6.2336*** |
| Law | 3.6917*** | 3.2842*** |
| Education | 4.3579*** | 2.4162*** |
| Arts | 15.3943*** | 9.8837*** |
| Languages | 9.1920*** | 5.2738*** |
| Political and social sciences | 9.5990*** | 5.5167*** |
| Psychology | 9.8709*** | 5.9034*** |
| Mathematics and physics | 2.5189*** | 2.8105*** |

| | | |
|---|---|---|
| Study abroad. Default: No study experience abroad | | |
| Erasmus experience | 0.9245 | 0.8227* |
| Other study experiences abroad | 1.205 | 1.0545 |
| Missing observation | 0.9828 | 0.8749* |
| Post-graduate studies | | |
| Training, apprenticeship aimed at gaining access to a liberal profession | 0.5342*** | 0.5280*** |
| Doctoral studies | 0.7701 | 0.8794 |
| Specialisation school | 0.4980*** | 0.4930*** |
| 1st level master degree | 0.7819** | 0.9925 |
| 2nd level master degree | 0.6551*** | 0.7579*** |
| Other type of master degree | 0.9797 | 1.0055 |
| Internship/work grant/on-the-job training | 0.8862 | 1.0935 |
| Public off-the-job training scheme | 0.9378 | 0.9424 |
| Study scholarship | 0.8107 | 0.7053** |
| Voluntary civil service | 1.2048 | 1.149 |
| Movers and stayers | | |
| Default: S(he) has not moved from the south and islands | | |
| Not moved from the north-western regions | 0.7283*** | 0.7428*** |
| Not moved from the north-eastern regions | 0.7059*** | 0.7436*** |
| Not moved from the central regions | 0.9053 | 1.0346 |
| Moved within the northern regions | 0.6776** | 0.8608 |
| Moved within the central regions | (omitted) | (omitted) |
| Moved within the southern and islands regions | 1.1731 | 2.0061*** |
| Moved to the north-west regions | 0.6593*** | 0.9224 |
| Moved to the north-eastern regions | 0.7215** | 0.9144 |
| Moved to the central regions | 0.8151 | 0.9153 |
| Moved to the southern and islands regions | 1.7623 | 1.1552 |
| Moved abroad | 0.6740** | 0.7722 |
| Constant | 0.0103*** | 0.0182*** |
| Number of observations | 17387 | 17387 |
| Pseudo $R^2$ | 0.1066 | 0.0989 |
| Correctly classified cases | 92.01% | 88.57% |
| Area under the ROC curve | 0.7493 | 0.7315 |

*Note: Legend: * p<0.05; ** p<0.01; *** p<0.001.*

*The figures in the Table represent odds ratios. The odds ratio associated with a characteristic j is the relative risk of being overeducated for individuals with a given characteristics with respect to the reference or default group. E.g., if the estimated odds ratio of, say, being IV years late in obtaining the degree equals 1.5, the corresponding group of graduates have a 50% higher probability of experiencing overeducation than the reference group of graduates who graduated in time. If the odds ratio equals 0.5 the individual with characteristics j have 50% lower probability of experiencing overeducation than the reference group.*

*Source: own elaboration of AlmaLaurea data.*

## Table 3. The wage penalty of overeducation and overskilling

| | *Overeducation (to get)* | *Overskilling (to do)* |
|---|---|---|
| ***Dependent variable:*** ***Natural logarithm of net monthly wages*** | ***(1)*** | ***(2)*** |
| ***Wage levels and wage penalty*** | | |
| ***In € (ratio to the well-matched income of € 1,299.91)*** | ***1075.48 (82.7%)*** | ***1122.24 (86.3%)*** |
| ***In log earnings*** | ***6.87*** | ***6.92*** |
| ***Unconditional estimates*** | | |
| ***OLS*** | -0.2081*** | -0.1568*** |
| ***Interval regression*** | -0.2463*** | -0.2088*** |
| ***Conditional estimates (without human capital variables)*** | | |
| ***OLS*** | -0.1264*** | -0.0697*** |
| ***Interval regression*** | -0.1414*** | -0.1003*** |
| ***Conditional estimates (with human capital variables)*** | | |
| ***OLS*** | -0.1220*** | -0.0692*** |
| ***Interval regression*** | -0.1319*** | -0.0967*** |
| ***Number of observations*** | 16591 | 16591 |
| ***Controlling for sample selection bias*** | Without instrumental variables | |
| ***Heckman model (ML simultaneous)*** | -0.1335*** | -0.0758*** |
| ***Heckman model (two steps)*** | -0.1336*** | -0.0758*** |
| | With instrumental variables (parents' education) | |
| ***Heckman model (ML simultaneous)*** | -0.1225*** | -0.0758*** |
| ***Heckman model (two steps)*** | -0.1337*** | -0.0759*** |
| ***Number of observations*** | 21605 | 21605 |

*Note: The table reports only the coefficients of interest. The OLS conditional estimates are obtained with all the control variables included in Table 2. The Heckit based on Maximum Likelihood simultaneous estimates are obtained with all the control variables included in Table 5.*
*Legend: * p<0.05; ** p<0.01; *** p<0.001.*
*Source: own elaboration of AlmaLaurea data.*

## Table 4. Earnings equations with Heckman correction for sample selection bias

| | Main equation | |
|---|---|---|
| Dependent Variable: | Log of net monthly wage | |
| Overeducation (to get) | -0.1225*** | |
| Overskilling (to do) | | -0.0758*** |
| Gender. Default: Men | | |
| Woman | -0.1547*** | -0.1834*** |
| Civil Status. Default: Single | | |
| Non-single | 0.1344*** | 0.1211*** |
| Civil Status. Default: Single | | |
| Non-single | 0.0557*** | 0.0225* |
| Number of children, Men | 0.1336*** | 0.1321*** |
| Number of children, Women | 0.0870*** | 0.0665*** |
| Non Italian | 0.0321 | 0.0041 |
| Type of high school diploma. Default: Science-based grammar school | | |
| Classical grammar school | 0.0152 | -0.0078 |
| Specialisation in teacher training | 0.0504*** | 0.0497*** |
| Language high school | 0.0119 | 0.0036 |
| Art school | 0.0007 | -0.0481 |
| Technical school | -0.0138 | -0.0133 |
| Professional school | 0.0505* | -0.0124 |
| Other high school diploma | 0.0204 | -0.0194 |
| Final grade at university. Default: Magna cum laude | | |
| 66-90 out of 110 | -0.0431** | -0.0635*** |
| 91-100 out of 110 | -0.0564*** | -0.0469*** |
| 101-105 out of 110 | -0.0373*** | -0.0207* |
| 106-110 out of 110 | -0.0493*** | -0.0197* |
| Field of study. Default: Engineering | | |
| Agriculture | -0.3152*** | -0.2617*** |
| Architecture | -0.2924*** | -0.2403*** |
| Economics and statistics | -0.1854*** | -0.0956*** |
| Physical education | -0.4398*** | -0.4149*** |
| Geology and biology | -0.1751*** | -0.1987*** |
| Law | -0.2893*** | -0.2369*** |
| Education | -0.3832*** | -0.2995*** |
| Arts | -0.3365*** | -0.3680*** |
| Languages | -0.2613*** | -0.2692*** |
| Political and social sciences | -0.2562*** | -0.1929*** |
| Mathematics and physics | -0.4767*** | -0.3683*** |
| Time spent obtaining a degree. Default: Curricular years | | |
| 1 extra-curricular year late | -0.1032*** | -0.0133 |
| 2 extra-curricular years late | -0.1434*** | -0.0441*** |
| 3 extra-curricular years late | -0.1536*** | -0.0481*** |
| 4 extra-curricular years late | -0.1692*** | -0.0792*** |
| 5 extra-curricular years late | -0.1670*** | -0.0941*** |
| Study abroad. Default: No study experience abroad | | |
| Erasmus experience | 0.0571*** | 0.0696*** |
| Other study experience abroad | 0.0617*** | 0.0648*** |
| No work experience | 0.0377*** | 0.0536*** |
| Post-graduate studies or professional experience | | |
| Training, apprenticeship aimed at gaining access to a liberal profession | -0.0674*** | -0.1375*** |
| Doctoral studies | 0.1143*** | -0.0622** |
| Specialisation school | 0.0186 | -0.0188 |
| 1st level master degree | -0.0213 | -0.0016 |
| 2nd level master degree | 0.0065 | 0.0105 |
| Other type of naster degree | 0.0207 | 0.0168 |
| Internship/work grant/on-the-job training | -0.0112 | 0.0252*** |

| | | |
|---|---|---|
| Public off-the-job training scheme | -0.0494*** | -0.0640*** |
| Study scholarship | 0.0239 | 0.0047 |
| Voluntary civil service | -0.1008*** | -0.1285*** |
| Stayer versus movers. Default: S(he) has not moved from the south and islands | | |
| Not moved from the north-western regions | -0.0994*** | 0.0696*** |
| Not moved from the north-eastern regions | -0.1146*** | 0.0626*** |
| Constant | 7.7484*** | 7.4408*** |
| Selection equation | | |
| Dependent Variable: | Being employed | |
| Gender. Default: Men | | |
| Woman | -0.1401*** | -0.1369*** |
| Civil Status. Default: Single | | |
| Non-single | -0.0496 | -0.0671 |
| Civil Status. Default: Single | | |
| Non-single | -0.1201*** | -0.1592*** |
| Number of children, Men | 0.1031 | 0.1207 |
| Number of children, Women | -0.0968* | -0.0895 |
| Non Italian | -0.0058 | -0.0259 |
| Type of high school diploma. Default: science-based grammar school | | |
| Classical grammar school | -0.0656** | -0.0575** |
| Specialisation in teacher training | -0.047 | -0.0002 |
| Language high school | 0.0061 | 0.0082 |
| Art school | -0.1476* | -0.1979** |
| Technical school | 0.006 | -0.0299 |
| Professional school | -0.1499** | -0.1083 |
| Other high school diploma | -0.1652** | -0.1785** |
| Final grade at university. Default: Magna cum laude | | |
| 66-90 out of 110 | -0.1151** | -0.2063*** |
| 91-100 out of 110 | -0.0036 | -0.0472 |
| 101-105 out of 110 | 0.0233 | -0.0065 |
| 106-110 out of 110 | 0.0788*** | 0.0705** |
| Field of study. Default: Engineering | | |
| Agriculture | 0.1248* | -0.2952*** |
| Architecture | 0.1490*** | -0.2296*** |
| Economics and statistics | 0.2108*** | -0.0308 |
| Physical education | -0.0112 | -0.5050*** |
| Geology and biology | -0.2395*** | -0.6327*** |
| Law | 0.1018*** | -0.4156*** |
| Education | 0.1144** | -0.1760*** |
| Arts | -0.2256*** | -0.6716*** |
| Languages | -0.1730*** | -0.5688*** |
| Political and social sciences | 0.1323*** | -0.1980*** |
| Mathematics and physics | 0.2920*** | -0.2277*** |
| Time spent obtaining a degree. Default: Curricular years | | |
| 1 extra-curricular year late | 0.1930*** | -0.0315 |
| 2 extra-curricular years late | 0.2321*** | -0.0038 |
| 3 extra-curricular years late | 0.2501*** | 0.0213 |
| 4 extra-curricular years late | 0.1874*** | -0.0639 |
| 5 extra-curricular years late | 0.1679*** | -0.1155** |
| Study abroad. Default: No study experience abroad | | |
| Erasmus experience | 0.1189*** | 0.0949** |
| Other study experiences abroad | 0.1059** | 0.1012* |
| No work experience | 0.0586 | 0.0879 |
| Post-graduate studies or professional experience | | |
| Training, apprenticeship aimed at gaining access to a liberal profession | -0.0398 | -0.0151 |
| Doctoral studies | -0.5771*** | -0.8161*** |
| Specialisation school | 0.0531* | 0.0766** |
| 1st level master degree | 0.0389 | 0.0408 |

| | | |
|---|---|---|
| 2nd level master degree | 0.0024 | -0.0212 |
| Other type of master degree | 0.041 | -0.0055 |
| Internship/work grant/on-the-job training | 0.1053*** | 0.1194*** |
| Public off-the-job training scheme | -0.0466 | -0.0670** |
| Study scholarship | -0.0899* | -0.1359** |
| Voluntary civil service | -0.1148** | -0.1471*** |
| Stayer versus movers. Default: S(he) has not moved from the south and islands | | |
| Not moved from the north-western regions | 0.9011*** | 1.1538*** |
| Not moved from the north-eastern regions | 0.9193*** | 1.2082*** |
| Father's education. Default: Compulsory or below | | |
| Secondary school | 0.0395* | 0.0354 |
| University | 0.0298 | 0.0441 |
| Don't know | -0.0395 | -0.041 |
| Mother's education. Default: Compulsory or below | | |
| Secondary school | 0.0781*** | 0.0614** |
| University | 0.0577* | 0.0209 |
| Don't know | 0.1400* | 0.1087 |
| Constant | 0.3824*** | 1.0523*** |
| Arthro | -1.3342*** | -0.0203* |
| Lnsigma | -0.6361*** | -0.8066*** |
| Number of graduates who declare their wage | 16591 | 16591 |
| Total number of graduates | 21605 | 21605 |

*Note: Pure pre-reform graduates in 2005, observed 5 years after graduation. ML simultaneous estimate.*
** p<0.05; ** p<0.01; *** p<0.001.*
*Source: own elaboration of AlmaLaurea data.*

**Table 5. Different specifications of the wage penalty**

| | OLS unconditional estimates | OLS conditional estimates | IR unconditional estimates | IR conditional estimates | Heckit **ML** (no instruments) | Heckit **ML** (with instruments) |
|---|---|---|---|---|---|---|
| **Model (1)** | | | | | | |
| **Well-matched (average wage: € 1299.91)** | | | | | | |
| Overeducation only (to get; € 1140.19) | -0.1629*** | -0.1101*** | -0.1832*** | -0.1126*** | -0.1182*** | -0.1207*** |
| Overskilling only (to do; € 1190.84) | -0.0935*** | -0.0179 | -0.1470*** | -0.0517*** | -0.0193 | -0.0230*** |
| Overeducation and overskilling **(genuine overeducation**; € 1044.01**)** | -0.2392*** | -0.1305*** | -0.2910*** | -0.1603*** | -0.1440*** | -0.1491*** |
| **Model (2)** | | | | | | |
| **Overskilling (definition with answers 2 and 3)** | -0.0643*** | -0.0293*** | -0.1158*** | -0.0544*** | -0.0330*** | -0.0330*** |
| **Model (3)** | | | | | | |
| **Overeducation** | | -0.1252*** | | -0.1341*** | -0.1331*** | -0.1333*** |
| **Overskilling** | | -0.0697*** | | -0.0970*** | -0.0742*** | -0.0743*** |

*Note: The table reports only the coefficients of interest. The OLS and IR estimates include the same control variables as in previous estimates. The Heckit estimates with no instruments are the same as in Table 4. Those with instruments use the educational level of parents as instrumental variables.*

*Model (1) refers to interactions of overeducation and overskilling. Model (2) refers to earnings equations relaive to a broader definition of overskilling. Model (3) has the same specification as in Table 3, except for the inclusion of the local unemployment rate of the destination areas for university graduates in OLS and IR estimates and the local unemployment rate of the area of residence for the Heckit estimates.*

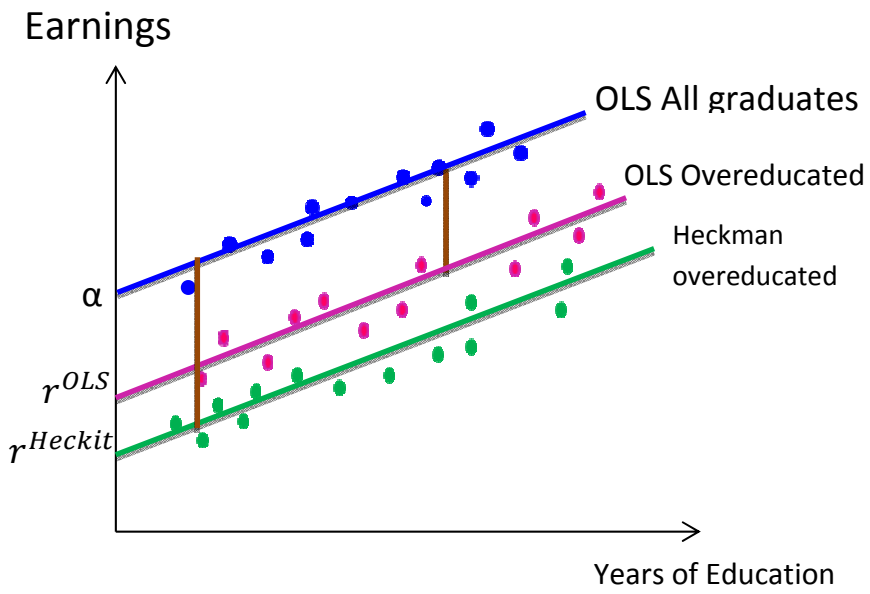
*Legend: * p<0.05; ** p<0.01; *** p<0.001.*

*Source: own elaboration of AlmaLaurea data.*

**Figures**

**Figure 1. Heckman correction of the wage effect of overeducation/overskilling**

**Panel (a): Job competition, job assignment and human capital model**




**Panel (b): Job search model**

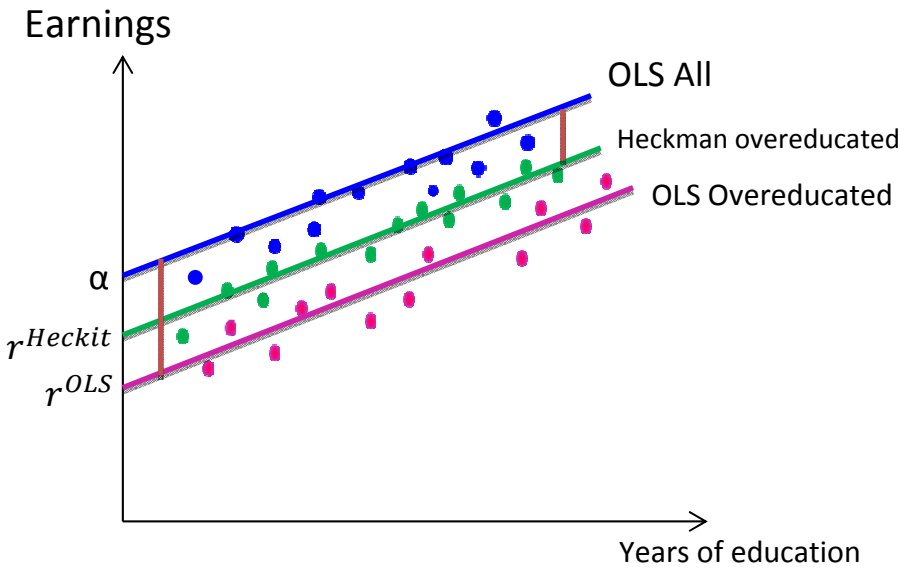
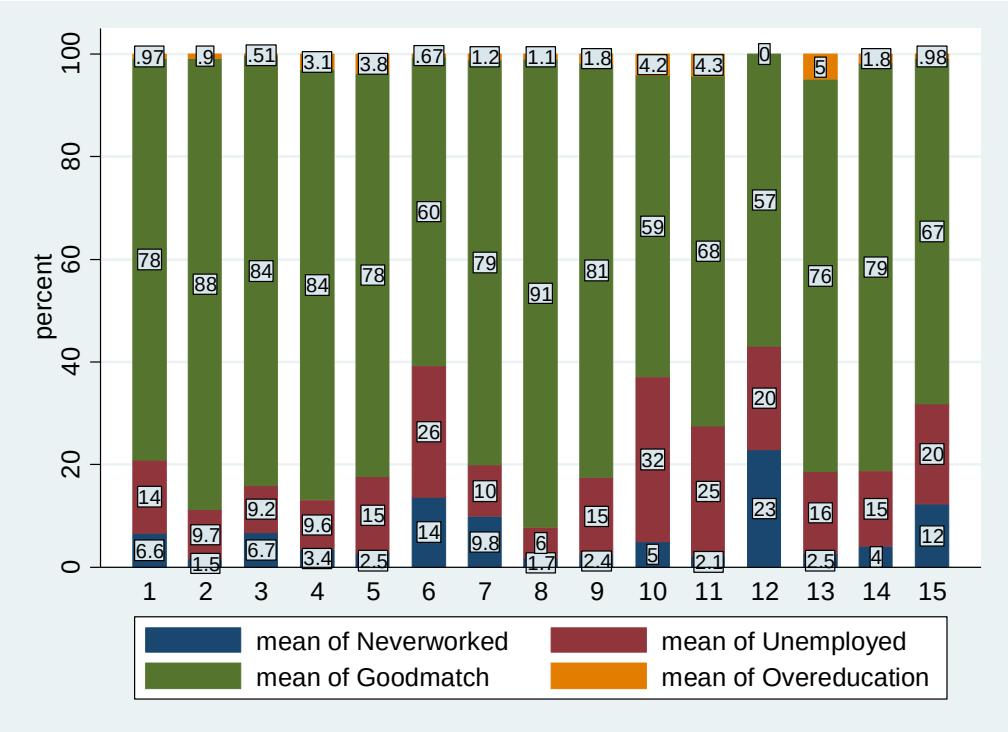
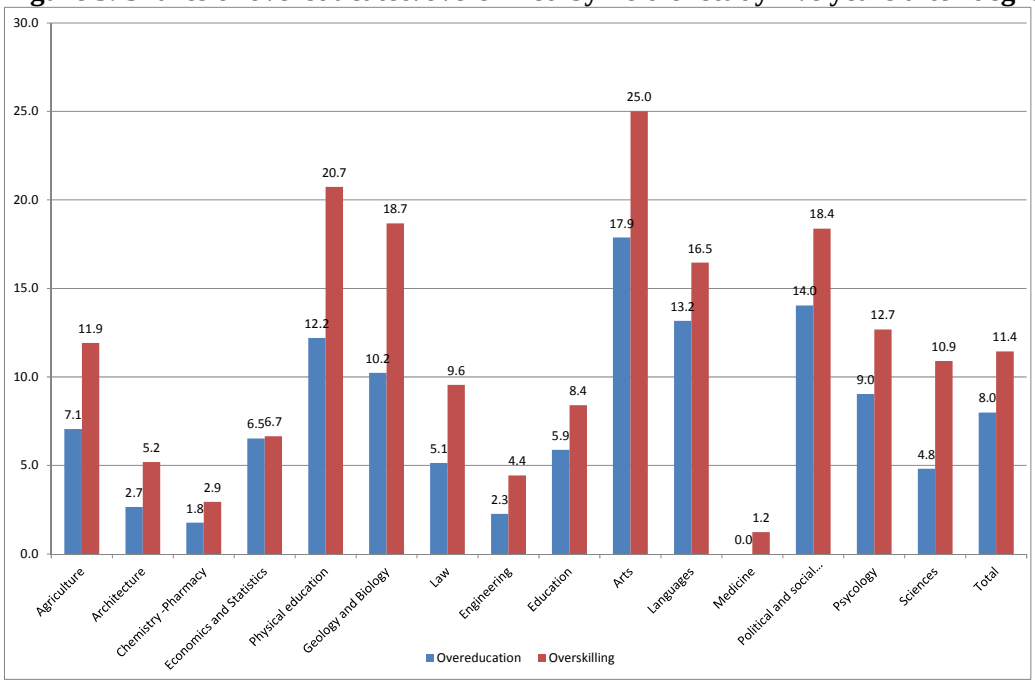


**Figure 2. The disruption of human capital by field of study (five years after graduation)**



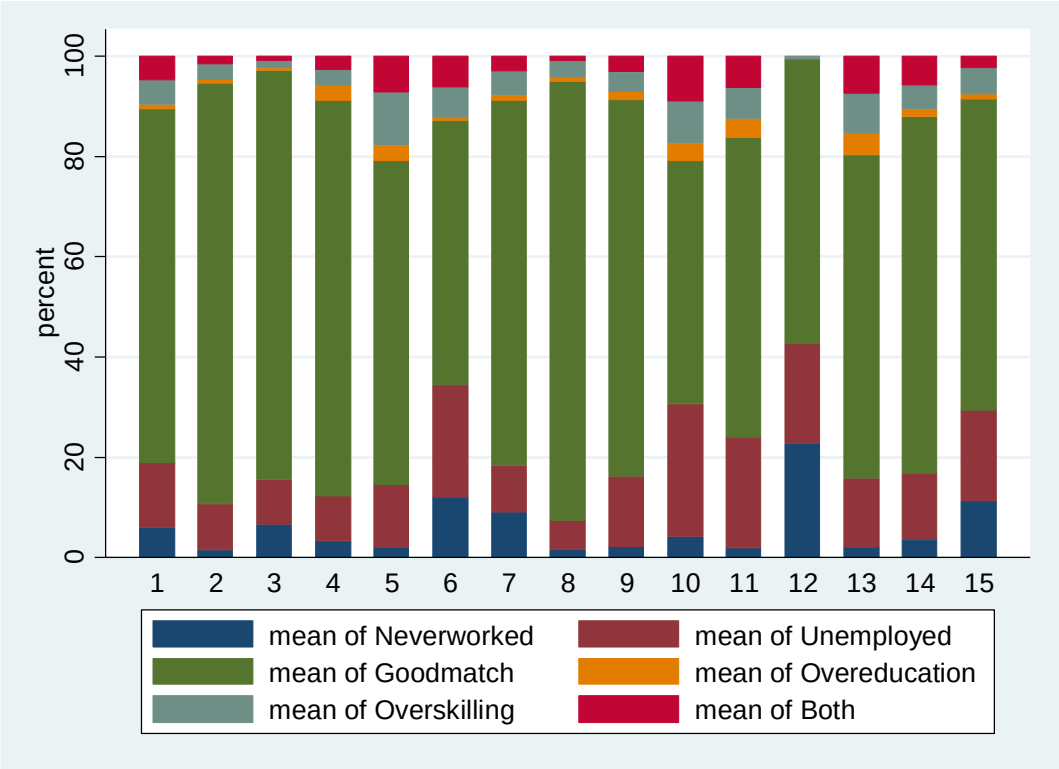

*Note: The shares are computed based on the total number of graduates.*
*1=Agriculture; 2=Architecture; 3:Chemistry-Pharmacy; 4=Economics and Statistics; 5=Physical education; 6=Geology and Biology; 7=Law; 8=Engineering; 9=Education; 10=Arts; 11=Languages; 12=Medicine; 13=Political and social sciences; 14=Psycology; 15=Sciences.*
*Source: own elaboration of AlmaLaurea data.*

**Figure 3. Shares of overeducated/overskilled by field of study five years after degree**




*Note: The shares are computed based on the number of employed graduates.*
*Source: own elaboration of AlmaLaurea data.*

**Figure 4. Genuine overeducation by field of study (five years after graduation)**

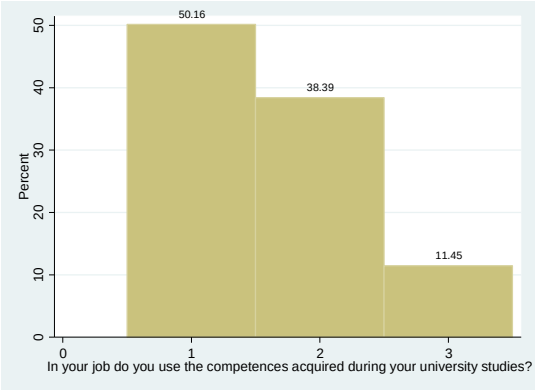
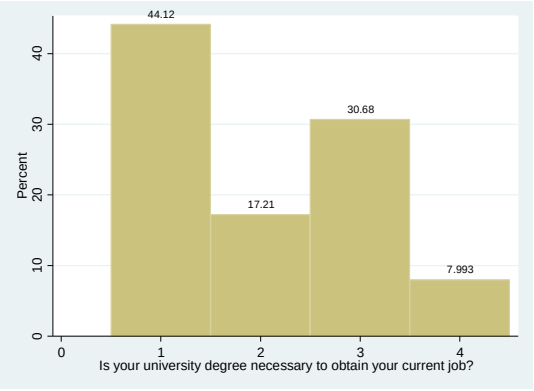



*Note: The shares are computed based on the total number of graduates.*

*1=Agriculture; 2=Architecture; 3:Chemistry-Pharmacy; 4=Economics and Statistics; 5=Physical education; 6=Geology and Biology; 7=Law; 8=Engineering; 9=Education; 10=Arts; 11=Languages; 12=Medicine; 13=Political and social sciences; 14=Psycology; 15=Sciences.*

*Source: own elaboration of AlmaLaurea data.*

Figure 5. Distribution of answers to the questions A16 and A17

| Panel (a): Overskilling (Question A16) | Pnel (b): Overeducation (Question A17) |
|---|---|
|  |  |





Note: The answers to the questions A16 and A17 are given in the data section.

*Source: own elaboration of AlmaLaurea data.*