Educational Mismatch of Graduates: a Multidimensional and Fuzzy Approach

Gianni Betti, Antonella D'Agostino, Laura Neri

Working Paper n. 71, June 2007

Educational Mismatch of Graduates: a Multidimensional and Fuzzy Approach[1]

Gianni Betti*, Antonella D'Agostino, Laura Neri***

* Department of Quantitative Methods, University of Siena, Italy
** Department of Statistics and Mathematics, University of Naples "Parthenope", Italy

Abstract

The present paper proposes a new statistical methodology for measuring the educational mismatch, seen as a problem of overeducation. It is hypothesized that educational mismatch is a latent variable which is difficult to measure because many factors converge to its definition; for this reason, multidimensional and fuzzy indicators are defined. They overcome the rigid definitions presented in literature by using a degree of membership to the set of overeducated workers. These new indicators are then used in order to explore the relationship between overeducation and different structural variables linked to the workers and the labor market. Our findings help to evaluate which factors have an effective impact in terms of educational match. Additionally, university reform introduced in the academic year 2001-2002 in the Italian higher education system is shown as not contributing to a reduction of the overeducation phenomenon.

Keywords: *Educational mismatch; Multidimensional indicators; Fuzzy sets, Factor analysis.*

[1] The work has been co-financed by PAR funds of the University of Siena for the project "Valutazione e meta-valutazione del sistema universitario: metodi, indicatori e modelli statistici". The paper is the result of the common work of all the authors; in particular G. Betti has written Sections 3 and 5.2; A. D'Agostino has written Sections 1, 2 and 6; L. Neri has written Sections 4 and 5.1.

1. Introduction

In the common sense, educational mismatch is defined as the lack of coherence between the required and offered educational level for a given job. The logical consequence of this issue is the presence of overeducated workers when the skills they bring to their jobs exceed the skills required for that job, and undereducated workers when these skills are inferior compared to those required for the job. Potentially, both these situations can have negative consequences for the labour market; anyway in recent decades there was a tendency towards a large increase of overeducated workers, and for this reason in the literature and in this paper we refer to 'educational mismatch', only the presence of overeducated workers. This phenomenon has been studied in several empirical works, which show that this is a problem that afflicts many industrialized countries (Alba-Ramirez, 1993; Groot, 1996; Allen and van der Velden, 2001; Cabral Veieira, 2005).

Recently, many empirical analyses have also been conducted in order to measure educational mismatch among graduates (Battu *et al.*, 1999; Belfield and Harris, 2002; Di Pietro and Cutillo, 2006). The social and economic repercussions of this issue can be very heavy, for this reason empirical analyses have also investigated on the returns of overeducation on earnings and on the job search (Cohn, 1994; Hartog, 2000; Dolton and Vignoles, 2000; Rubb, 2003; Di Pietro and Urwin, 2006).

In all above-mentioned papers, a definition of educational mismatch has been used in order to carry out the empirical analysis. These definitions can be classified into four different approaches that are divided into "objective" and "subjective" measures (Groot and van den Drink, 2000). The objective measures can be based on: i) the comparison between the actual educational level and the job level requirements; ii) the comparison between the years of education attained and the average educational level within the occupation of the workers.The subjective measures are based on self-assessment of the worker by: iii) a direct question whether they are overeducated or undereducated for the work they do, and iv) a question on what the minimum educational level required for the job they have is.

These four definitions have a common structure as they are based on a unidimensional concept of the educational mismatch; this is measured by a dichotomous variable that divides the collectivity of the workers into two groups: all those who have positive mismatch between educational level attained and their job (overeducated) and all the other. Anyway, the concept of educational mismatch is a very complex issue in which many causes converge and a definition based on a simple univariate measure can be too heavy a simplification of the phenomenon. Moreover, the dichotomous definition into two groups of workers can at the same time present a very misleading picture of the reality, as different degrees of membership of overeducation may exist which this simple indicator cannot take into account.

These considerations form the basis of this paper. We propose, in fact, a multidimensional approach for defining the educational mismatch and, at the same time, we use the fuzzy set theory (Zadeh, 1965; Dubois and Prade, 1980) in order to create fuzzy indicators overcoming the division of workers into two groups. The new measures define a degree of membership to the set of overeducated workers.

The proposed mismatch indicators are then used in order to explore the relationship between overeducation and different structural variables linked to the workers and to the labor market characteristics. Among our aims, some ideas and theories about the possible impact of the recent reform of the Italian higher education system (introduced in the academic year 2001-2002) on educational mismatch are also included.

The structure of the paper is as follows. After the present introduction, in Section 2 the methodological framework is briefly introduced. Then in Section 3 the data used in the empirical analysis are described, and the latent dimensions of educational mismatch are identified in Section 4. In Section 5 the results obtained from the empirical analysis are discussed. Some final remarks conclude the paper in Section 6.

2. Methodological framework

In literature, different approaches have been proposed in order to define and to measure the educational mismatch of graduated workers. The common structure of these approaches (that we label as 'traditional' approach) is based on a unidimensional and dichotomous indicator of the educational mismatch, so that in the analysis the collectivity is divided into two groups: the overeducated group and the not overeducated group. Actually, it is very difficult to measure the educational mismatch of graduated workers, as many observable and unobservable factors can influence it. For this reason, we propose measuring the educational mismatch using a multidimensional and fuzzy approach. The multidimensional nature of such an approach implies that different indicators are used in order to define the mismatch and the fuzzy nature helps us to overcome the rigid division of graduated workers into two groups. In other words, we define a degree of membership to the set of overeducated workers.

Following this perspective, we suppose that a latent variable ξ exists which identifies the educational mismatch. As ξ is not directly observable, instead we use the K observable variables X_i ($i=1...K$) that can be used in order to obtain a multidimensional measure of educational mismatch. Of course, the set of variables chosen and consequently the number of possible variables (K) that can be used in the analysis depends on available data and on the choices of the researcher.

Among these variables, the traditional indicators, used in many empirical analysis, can be included as well as other potential explicative factors such as a measure of the personal satisfaction for the job they have (in this case the hypothesis is that a worker is more satisfied with his/her job if he/she is not overeducated) and other variables measuring the personal satisfaction concerning earnings in comparison with the educational level attained. In general, variables which contain useful information for the definition of educational mismatch can have an ordinal scale with more than two categories. To treat them as a metric, we propose to assign scores to each category of each variable, following the approach proposed by Cerioli and Zani (1990). In this way, for each variable *i* with ordered categories 1 (least mismatched) to, say, M (most mismatched), graduates in a generic category *m* will have a score given by:

$$S_i = s_{(m)} = \frac{m-1}{M-1} \qquad (1)$$

so that the computed score will range between 0, for the least mismatched, to 1, for the most mismatched.

Chosen the set of the observable variables and having transformed the original items as metric values, the following methodological step is to discover if a number of potential dimensions H exists which is related to the multidimensional concept of the educational mismatch. In other words, the aim is to discover if the observable variables can be grouped and therefore if they can compose a restricted number of dimensions *H* (*H*<*K*). We would expect these dimensions to be easily interpretable as, for example, the earnings dimension that take into account only the variables concerning earnings. An explorative factor analysis can be used to identify the latent dimensions and then a confirmative factor analysis can be applied in order to check the adaptability of the identified model to the observed data (for instance, see Whelan *et al.*, 2001, in the field of poverty and social exclusion).

In factor analysis, the latent variables are traditionally called factors. Let ξ_j *(j=1…H)* be the latent factors discovered, we will be able to state which are the observable variables that give contribution to explain the variability of each factor; in this sense, the aggregation process of the single indicators into the multidimensional measure can be described by a weighted $I(.)_w$ or simple mean $I(.)$ into each dimension ξ_j:

$$I_{m.f.}\left(\xi_j\right)_w = \frac{\sum_{i=1}^{K\in\xi_j} w_i S_i}{\sum_{i=1}^{K\in\xi_j} w_i} \tag{2}$$

if $w_i = 1$ we obtain the simple mean. The weights w_i are strictly related to the particular framework of analysis[2]. In our analysis we propose using the factor loading of each variable as weight functions, computed by the confirmative factor analysis.

[2] In poverty analysis different weight functions have been specified. All these specifications respect two main principles. Firstly, the weights should be determined by the variable's power to 'discriminate' among individuals in the population, that is, by its dispersion. Secondly, from a non-redundant point of view, it is necessary to limit the influence of those variables that are highly correlated with the others included in the analysis. For a further discussion of the issue in the context of poverty measurement, see, among others, Betti and Verma (1999), Filippone *et al.* (2001), Betti *et al.* (2006) and Aassve *et al.* (forthcoming).

3. The AlmaLaurea Survey of Italian Graduate Employment

The measures of educational mismatch proposed in the present paper are applied to the Survey of Italian Graduate Employment conducted by AlmaLaurea, a consortium of 36 Italian universities. Every year the Survey contacts graduate students at one year, three years and five years from graduation; only students who graduated in the summer sessions are interviewed. The set of interviewees is therefore not a sample of the population of graduates, but just a sub-census of it.

As agreed by AlmaLaurea and the member universities, the results regarding the graduates' employment and postgraduate careers are delivered to each university in the following three distinct timeframes (Cammelli, 2007):

- in December each year, the initial results of the recently completed survey are delivered to the university governing bodies;

- the final results regarding graduate entry into the job market are published in February; these refer to the overall group of graduates examined and are broken down by degree courses. The results collected in this Report and available on the Internet are presented during a specially held national conference;

- in the summer, each university is sent the processed data concerning their own graduates with in-depth studies and analyses with breakdown charts by faculty. The relevant microdata is also made available to the universities, allowing them to analyse the survey in more detail and to trace the results back to the original degree courses/categories.

The survey of the employment careers of graduates is now in its ninth edition; microdata are currently available up to the eighth edition only. The research project, which has been running for some years now (all the documentation is available on

Internet at *www.almalaurea.it*) aims to investigate the employment and further training careers of graduates in the first five years after receiving their degrees.

The survey of the eighth edition was conducted between September and November 2005 and targeted graduates who had received their degrees in the summer sessions of 2004, 2002 and 2000. In December 2005, just a few weeks after the survey was concluded, the first aggregated results were sent to the universities participating in the survey. In particular in the present paper we make reference to the graduates at the University of Siena in the summer sessions of 2004 and interviewed at one year from graduation during the eighth edition of the Survey; the corresponding microdata were released during summer 2006.

To overcome the fact that the AlmaLaurea Survey interviews graduates in the summer sessions and the fact that some of the graduates have not participated in the survey (non-response), the sample of graduates has been subjected to the re-proportioning procedure that is specifically applied in such cases. The re-proportioning procedure, also known as calibration[3], has been performed on the basis of some variables related to the phenomenon under investigation, and known for the total population (we make reference to the AlmaLaurea Survey on Graduates Profile, year 2004). We have chosen six variable distributions: (i) gender, (ii) age categories at graduation, (iii) level of degree (pre-reform, 4 years or more or post-reform, 3 years), (iv) degree course grouping, (v) area of residence at graduation and (vi) graduation mark in classes.

[3] Calibration is done through an iterative procedure that attributes a "weight" to every graduate interviewed so that the relative distributions of the re-proportioned variables are as similar as possible to those observed across the entire graduate population at the University of Siena during year 2004. A graduate whose characteristics are very common among the population but not in the AlmaLaurea sample, will be attributed a proportionally higher weight. Conversely, a graduate with characteristics that are common in the AlmaLaurea sample but not across the whole population will be attributed a proportionally lower weight.

A small proportion of the interviewed graduates have not responded to some questions (item non-response); in order to avoid eliminating those individuals from the sample, we have decided to impute missing data. The imputation procedures used here are based on the "sequential regression multivariate imputation" (SRMI) approach adopted by the imputation software (IVE-ware). The method proposed by the authors of the software (Raghunathan *et al*., 2001) constructs the imputed values by fitting a sequence of regression models and drawing values from the corresponding predictive distribution, under the hypothesis of Missing at Random (MAR) mechanism. The procedure is a variant of the estimation-maximisation (EM) algorithm and follows a Bayesian paradigm. The sequential multivariate model is used for more complete imputation of the variables, which at the same time can safeguard their variance and their inter-correlation.

4. Identification of the latent dimensions

The reference data of the analysis regards 665 individuals. The 665 units represent the subgroup individuals that had a job and that gave a complete interview to the AlmaLaurea survey one year after the degree. The following six observable variables X_i (i=1…6), connected with the "mismatch concept", have been used for the analysis:

X_1 = Use of expertise;

X_2 = Degree requirement for employment;

X_3 = Work satisfaction;

X_4 = Coherence between degree and job;

X_5 = 'Is the salary adequate for the degree reached?';

X_6 = 'Is the salary adequate for the job level?'.

As already specified in Section 2, the original variables X_i have been transformed into the corresponding metric variables S_i and then an explorative factor analysis[4] has been applied to identify the latent dimensions ξ.

In order to decide how many factors ξ_j *(j=1...H)* have to be retained, we referred to the Kaiser[5] criterion (Hakstian *et al.*, 1982), according to which only those factors whose eigenvalues are greater than 1 are retained.

Looking at the eigenvalues in Table 1, only two eigenvalues are greater than 1, therefore only two factors are retained. The factors retained explain about 67% of the total variance. In order to increase the interpretability of the new dimensions an orthogonal Varimax rotation (Kaiser, 1960) has been used.

Table 1. Exploratory factor analysis

Eigenvalue	**Difference**	**Proportion**	**Cumulative**
2.3951	0.7453	0.3992	0.3992
1.6499	0.8965	0.2750	0.6742
0.7534	0.2257	0.1256	0.7997
0.5277	0.1560	0.0880	0.8877
0.3717	0.0696	0.0620	0.9496
0.3022		0.0504	1.0000

Looking at the rotated factor loadings in Table 2, it can be observed which are the loadings that most strongly characterize each factor. ξ_1 effectively represents "the overeducation factor", since it is essentially characterized by the following variables: i)

[4] The explorative factor analysis has been conducted using the SAS FACTOR procedure. The correlation matrix among the six variables has been computed in order to be sure that a significant correlation structure exists among the variables.

[5] The Kaiser criterion accurately identifies the true number of factors if the sample size is greater than 250 units and the mean communality is greater than 0.6; in our situation, the criterion seems to be pertinent, since the sample size is 665 and the mean communality, expressed by the ratio between the total communality estimates (4.05) and the total number of variables considered (6) is about 0.67.

use of expertise; ii) degree requirement for employment; iii) work satisfaction; iv) coherence between degree and job. Otherwise, ξ_2 represents "the earning factor", since it is essentially characterized by two items: i) 'Is the salary adequate for the degree reached ?'; ii) 'Is the salary adequate for the job level ?'.

Table 2. Rotate Factor Matrix

Variables	**Factor Loading**		**Final communality estimates**
	ξ_1	ξ_2	
S_1	0.8851	-0.0038	0.6996
S_2	0.8362	-0.0212	0.5075
S_3	0.6968	-0.1482	0.7803
S_4	0.6306	0.2741	0.8013
S_5	-0.1174	0.8874	0.4727
S_6	0.1137	0.8760	0.7835
Explained variance	2.3925	1.6524	

With the explorative factor analysis, the latent structure based on the established set of variables has been identified; now the aim is to find evidence for the significance of the two identified factors using a confirmative factor analysis (CFA, Nasselroade and Baltes, 1984)[6]. In CFA each observed variable has an error term associated with it[7]; the error term is represented by δ. The equation specifying the relationship of the observed variables (in our application we refer to the transformed variables S_i) to the factors and the error term is the following:

[6] In the CFA, the researcher can specify a structure of the a priori factor models, according to the theories about how variables ought to be related to the factors.

[7] The error terms are similar to the residuals in a regression model, because they are part of each observed variable that is not explained by the factors. However in CFA the error terms contains also measurement error due to the lack of reliability of the observed variables.

$$S = \lambda\xi + \delta \quad (3)$$

where λ represents the factor loadings and ξ represents the factors themselves. The solutions for λ and δ cannot be obtained through standard regression methods as ξ is a latent factor. Therefore, the covariance matrix of the observed variables is used to find a solution for the element of the matrices λ and δ. Labelling with $\hat{\Sigma}$ the sample covariance matrix and with Σ the population one, it is necessary to find the relationship between Σ (or $\hat{\Sigma}$) and the elements of λ, ξ and δ. Overcoming the algebra (Bollen, 1989), the relationship is the following:

$$\Sigma = \lambda\Phi\lambda' + \theta\delta \quad (4)$$

where Φ is the matrix of correlations or covariances among the factors (ξ s) and $\theta\delta$ is the matrix of correlations or covariances among the error terms. The previous equation is solved so as to find values for elements λ, Φ and $\theta\delta$.

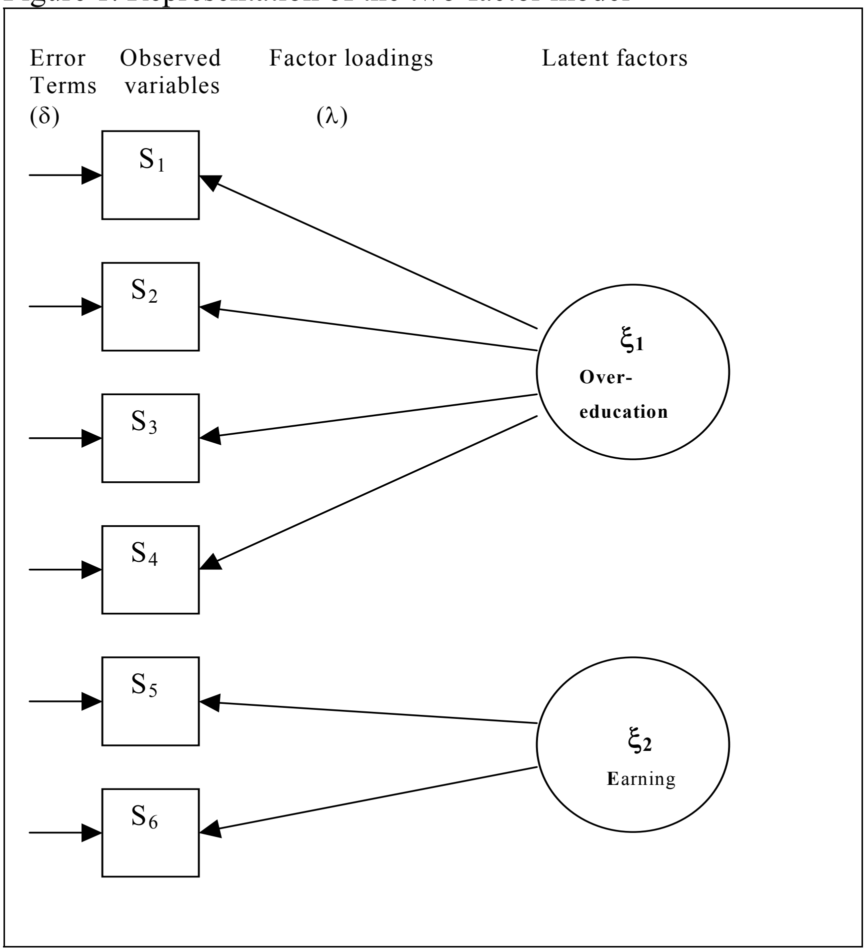
Now it is necessary to specify the structure of the matrices λ, Φ and $\theta\delta$, so as to identify which elements are to be included. The explorative analysis has provided results which are plausible and form the basis of our specification as follows; the λ matrix would be specified to include only the loadings of the items designated to measuring each factor: it means that six factor loadings are specified; assuming no correlation between factors, the Φ matrix of the correlation would be an identity matrix; finally one measurement error for each item would be estimated. Figure 1 gives a representation of the hypothesized two-factor structure.

Before taking into account the results of the statistical tests and of the overall fit indices, it is worth paying attention to the identification of the model[8]. In the

[8] Identification refers to whether the parameters of the model can be uniquely determined. Models that have more unknown parameters than pieces of information are called under-identified models, and cannot

hypothesized two-factor structure, the number of parameters to be estimated would be 12, corresponding to the 6 factor loadings plus the 6 error terms. The number of known parameters in the model corresponds to the 21 elements[9] in and below the diagonal of the variance-covariance matrix Σ. Thus in the application the model is over-identified, it can be solved uniquely and it can be statistically tested.

Figure 1. Representation of the two-factor model



be solved uniquely. Models with just as many unknown parameters as pieces of information are called just-identified models, and can be solved uniquely, but cannot be tested statistically. Models with more information than unknown parameters are referred to as over-identified models, these models can be solved uniquely and can be tested statistically.

[9] If p is the number of observed variables, the number of unique values in the covariance matrix is p(p+1)/2. That value is given by the number of terms below the diagonal plus the variance elements.

Table 3 shows the estimates of the factor loadings and the error terms for the specified model. Moreover, the t values are obtained and all of them show that the parameters of the model are statistically significant. However, it is evident that the items S_1 and S_4 have loadings that are greater than the other items for the same factor; this is an important result, showing that the effective use of expertise and the coherence between degree and job are the most important items for explaining the mismatch concept as far as the relationship between studies and job is concerned; in this sense it is enforced the label given to the factor ξ_1 that seems to express the effective, more than formal, overeducation concept.

Table 3. ML standardized estimates for factor loadings (λ) corresponding to t-statistics and error terms (δ)

S	=	λ	*	ξ	+	δ	t-value
S_1	=	0.749	*	ξ_1	+	0.66	20.1811
S_2	=	0.546	*	ξ_1	+	0.84	14.1176
S_3	=	0.512	*	ξ_1	+	0.86	13.1326
S_4	=	0.910	*	ξ_1	+	0.42	25.2550
S_5	=	0.650	*	ξ_2	+	0.76	18.6697
S_6	=	0.920	*	ξ_2	+	0.39	33.5387

It is worth noting that the traditional indicator adopted for measuring the mismatch concept, as formal overeducation, is the third out of four items explaining the factor ξ_1 (having $\lambda = 0.546$). On the other hand, also the factor relating to earnings ξ_2 seems to be very important, since both the factor loadings (related to S_5 and S_6) are high. The significance of the factor loadings is very important because it indicates that the items have significant loadings on the factors they were intended to measure. Moreover, there are no unreasonable parameter estimates, such as negative variance, and all the parameters have the expected signs. In order to test all model parameters simultaneously, it is necessary to observe some overall fit statistics[10] (see Table 4). The indices of goodness of fit, suggest that the model is quite good

A Chi Squared value of 75.4 with 9 degrees of freedom is clearly significant, indicating that the model does not adequately account for the observed covariance among variables. In such cases, it is reasonable to think that the significant chi-square is due to the large sample size, rather than to any serious misspecification of the model. Concluding, the inspection of the parameter values, the t statistics and the overall fit indices, support the hypothesized two-factor structure.

Table 4. Goodness of fit statistics

Goodness of Fit Index (GFI)	0.9635	RMSEA Estimate	0.1054
Parsimonious GFI	0.5781	RMSEA 90% Confidence Limit	0.0841-0.1280
Chi-Square	75.3387	Non-normed Index (NNFI)	0.9027
(DF, p-value)	(9; <0.0001)	NFI	0.9346

[10] The GFI (Joreskog and Sorbom, 1993) can be thought of as the amount of the overall variance and covariance in S that can be accounted for by $\Sigma(\theta)$ and is similar to the determination coefficient in the multiple regression model. Bentler and Bonnet (1980) introduced a class of fit indices named comparative fit indexes. These indexes compare the fit of the hypothesized model to a null model, in order to measure the amount by which the fit is improved. The two indexes are the normed fit index (NFI) and the non-normed fit index (NNFI). The first one ranges between 0 and 1 with higher values indicative of a greater improvement in fit; the NNFI does not ranges in 0-1, however values greater than 0.9 are typically considered as indicative of a good fit. Finally the root mean square error of approximation (RMSEA) that is essentially a measure of lack of fit for degree of freedom. According to their experience Browne and Cudeck (1993) suggested that RMSEA values of 0.05 or less indicate a close approximation and that values of up to 0.08 suggest a reasonable fit of the model in the population.

The goodness of fit of the identified structure has suggested experimenting a weighting system based on the factor loadings to compute the different multidimensional measure of the mismatch concept. In detail, the weighted indicator for each dimension is computed substituting the factor loadings obtained with the CFA into w_i in formula (2); in this way two multidimensional indicators are obtained: $I_{m.f.}(\xi_1)_w$ and $I_{m.f.}(\xi_2)_w$.

5. The analysis of educational mismatch

The aim of this section is to discover the substantial different information included in the traditional indicator with respect to the new multidimensional indicators, highlighting if the new indicators give some new evidence. In order to analyse the mismatch phenomenon from a different point of view, the indicators are computed disaggregating by a set of relevant collected variables, such as individual, education and job characteristics. However, in order to facilitate comparisons among the variety of indices computed (reported in Table 5), it is necessary to specify some preliminary remarks.

5.1 Preliminary remarks

The remarks regard essentially two points: i) the differences between the simple $I_{m.f.}(.)$ and the corresponding weighted multidimensional-fuzzy indicators $I_{m.f.}(.)_w$; ii) the differences between the new multidimensional indicators and the traditional one $I_t(.)$[11].

Considering the former aspect (i), from Table 5 it is evident that the results using the specification with the weights $I_{m.f.}(.)_w$, are substantially similar to those corresponding to the simple one $I_{m.f.}(.)$. The reason for this similarity is probably related to the low variability of the factor loadings across the items composing each factor (see Section 4). According to this evidence, the following discussion will be based only on the weighted indices. Considering the latter aspect (ii), it is very important to show that even if both measures (multidimensional-fuzzy and traditional) range between 0 and 1, the interpretation of the measures is not the same. The traditional measure represents a proportion, that is, $I_t(S_2) = 0.74$ means that 74% of the individual of the collectivity declares that the obtained degree was not required for the actual job (consequently for the remaining 26% of the individuals the degree was required in order to obtain the actual job). The multidimensional indicators are cardinal measures, representing a degree of membership to the set of overeducated workers; the maximum level of mismatch is equal to 1; a worker, in order to have the multidimensional indicator equal to 1 needs to have all the indicators included in the multidimensional measure equal to the maximum level of mismatch. In this sense, $I_{m.f.}(\xi_1)_w = 0.74$ means a relevant degree of mismatch, being quite close to the maximum level.

---

[11] $I_t(.)$ is the proportion of individuals answering "No" to the question "Is the degree required for employment ?" (X_2).

Table 5. Mean values of the mismatch indicators across structural variables

Structural variables	Weighted %	Mismatch indicators				
		$I_{m.f.}(\xi_1)$	$I_{m.f.}(\xi_2)$	$I_{m.f.}(\xi_1)_w$	$I_{m.f.}(\xi_2)_w$	$I_t(S_2)$
Gender						
Female (n=215)	58%	0.41	0.60	0.42	0.58	0.78
Male (n=152)	42%	0.35	0.47	0.35	0.45	0.70
Field of Education						
Economic (n=62)	19%	0.36	0.50	0.36	0.48	0.77
Pharmacy (n=17)	4%	0.18	0.45	0.18	0.43	0.15
Law (n=37)	11%	0.49	0.50	0.49	0.48	0.88
Engineering (n=24)	9%	0.35	0.48	0.35	0.47	0.88
Humanities (n=118)	30%	0.46	0.70	0.46	0.68	0.86
Medicine (n=35)	7%	0.15	0.43	0.15	0.42	0.01
Math. /Physical Sc. (n=22)	5%	0.44	0.51	0.46	0.50	0.62
Political Science (n=52)	18%	0.37	0.46	0.37	0.45	0.81
Employment sector						
Public (n=82)	22%	0.31	0.57	0.30	0.56	0.68
Private (n=285)	78%	0.41	0.54	0.41	0.52	0.76
Employment contract						
Self-employment (n=31)	7%	0.36	0.47	0.37	0.47	0.58
Permanent employment (n=97)	27%	0.41	0.48	0.41	0.46	0.81
Temporary employment (n=79)	21%	0.42	0.51	0.43	0.49	0.71
Collaboration (n=107)	30%	0.33	0.66	0.32	0.63	0.71
Without contract (n=25)	7%	0.49	0.61	0.49	0.59	0.73
Apprenticeship or trainee (n=28)	9%	0.38	0.49	0.39	0.47	0.85
Firm size						
1 (n=21)	5%	0.42	0.53	0.42	0.51	0.70
2-5 (n=77)	20%	0.35	0.64	0.35	0.63	0.62
6-14 (n=58)	16%	0.41	0.52	0.41	0.49	0.76
15-49 (n=61)	16%	0.42	0.66	0.42	0.64	0.79
50-99 (n=23)	6%	0.45	0.54	0.46	0.52	0.89
100 or more (n=127)	37%	0.37	0.46	0.37	0.45	0.77
Employment is						
same as before the degree (n=122)	34%	0.44	0.57	0.44	0.55	0.87
another job after the degree (n=54)	14%	0.45	0.57	0.46	0.54	0.83
the first job (n=191)	52%	0.33	0.53	0.33	0.51	0.63
Kind of university course						
Post reform (n=90)	34%	0.41	0.54	0.41	0.52	0.80
Pre reform (n=277)	66%	0.38	0.55	0.38	0.53	0.71
TOTAL SAMPLE (n=367)	100%	0.39	0.55	0.39	0.53	0.74
Medicine new degree (n=298)		0.23	0.77	0.23	0.77	0.33

Moreover, we point out that the comparison between the multidimensional indicators and the traditional one regards mainly the indicators $I_{m.f.}(\xi_1)_w$ and $I_t(S_2)$ as their meaning properly regards overeducation[12]. On the other hand, the multidimensional-fuzzy indicator regarding the earning factor is something new with

[12] In fact, $I_{m.f.}(\xi_1)_w$ includes information regarding job satisfaction, the use of expertise and the coherence between study and work, but also the measure on which the traditional indicator is based.

respect to the traditional one, adding new relevant information to the analysis of the mismatch concept. The different interpretation of the traditional and multidimensional indicators makes it reasonable for the values of the indicator $I_t(S_2)$ to be generally greater than $I_{m.f.}(\xi_1)_w$. The reason is likely to be related to the fact that the traditional indicator is based on a one-dimensional measure of mismatch, so that the individual is classified as belonging or not to the set of overeducated workers; whereas, the multidimensional-fuzzy indicator jointly takes into account, a set of indicators, so that an individual may present relevant aspects of the mismatch in one indicator and less relevant ones in another.

Before analyzing the figures in Table 5 from a substantial point of view, a couple of aspects regarding the analysis conducted by groups should be specified: a) the composition of the graduates in medicine; b) the size of some subgroups. With respect to aspect (a), it is relevant to note that 298 out of 333 graduates in medicine obtained the post reform degree and only the remaining 35 graduated with the pre reform degree. Knowing the situation of the Faculty of Medicine and checking some characteristics of the 298 individuals who graduated after the reform[13], it can clearly be deduced that they are paramedics (nutritionists, physiotherapists, obstetricians, nurses, etc.). It is known, in fact that in the academic year 2004-2005, many individuals with the old 'university diplomas'[14], sat some more exams and in this way obtained the first level university degree[15]. Considering the peculiarity of the above mentioned group we decide to present

[13] The 298 graduates in medicine after the reform present, on average, a higher age with respect to the remaining group of the post reform graduates; particularly the average age of the two groups is 39 and 28 years old; moreover most of them already had employment before the degree.

[14] The Italian law introduced in 1990 regarding 'university diplomas' (*Diplomi Universitari, DU*), that is to say courses lasting three years, although occasionally two years as well, clearly sought to provide a professional grounding, albeit not without a component of basic university instruction.

[15] In order to clarify the discussion, it is useful to briefly describe the higher educational system in Italy. The old university course was composed of only one level, that is to say universities conferred only one

results in Table 5 distinguishing that group, so that the results regard the 298 graduates in medicine after the reform (called "Medicine new degree"), and the remaining 367 individuals (called "Total sample" in Table 5). As far as aspect (b) above is concerned, we point out that differences between indicators will be considered only at point estimates level, because some groups of graduates are so small in size, that it does not make sense to test the significance of the differences between different figures. For each structural variable we also report the weighted percentage of the several modalities of the variable itself.

5.2 Educational mismatch in the University of Siena

Table 5 summarizes the effects of a set structural variables such as: gender, field of education, some characteristics of the current employment, kind of degree (pre-reform or post reform) for the different specifications of the indices described in Section 4. In almost all the cases, the relationship between overeducation and the other variables considered offers many ideas for an interesting discussion.

Considering the gender as disaggregation variable, we can immediately realize that the great majority of graduates are females (58%). However, even if females invest more in their human capital, they are not repaid in the labor market; in fact, according to both indicators $I_{m.f.}(\xi_1)_w$ and $I_t(S_2)$, they are more overeducated compared to males. Moreover, the most remarkable difference between male and female indicators concerns the earning factor $I_{m.f.}(\xi_2)_w$: this is equal to 0.45 for male and equal to 0.58 for

qualification, the degree, at the end of courses of study. The reform has introduced more levels in terms of university qualifications. University studies begin for all enrolled students with a degree course that lasts three years. The degree thus becomes the first-level university qualification. Then, the graduates can enter the world of work and leave universities for good, or they can continue their university studies at the second level immediately after the degree or during their working careers.

females. This result can be also explained by considering that predominantly female occupations pay less (Blau and Kahn, 2000), as noted in many empirical researches that attribute wage differential to discrimination in the labor market (Jarrell and Stanley, 2004).

How can our findings be explained with respect to employment variables (employment sector, employment contract, firm size, etc.)? Considering the employment sector of the graduates, both $I_{m.f.}(\xi_1)_w$ and $I_t(S_2)$ are remarkably lower for the public than for the private sector. The explanation of the results is probably related to the fact that in order to be employed in a public job the individuals have to participate in a public selection, for which specific degrees are usually requested. Concerning the earning indicator, our findings show a higher mismatch in the public sector than in the private one. It is quite difficult to interpret this result, as evidence for Italy suggests a positive wage differential between the public and the private sector (Lucifora and Meurs, 2006). Any way, this difference is very small (0.04) and probably due to sample data, much could be done to further improve our knowledge on this point.

Considering the employment contracts, we can observe that the size of some groups (without contract, apprenticeship or trainee, self employment), are quite small. Considering the other groups, with respect to the $I_{m.f.}(\xi_1)_w$ index it is reasonable that permanent or temporary employment contracts present higher values (respectively 0.41 and 0.43) than collaboration contract (0.32). This result maybe can be interpreted in the light of the Biagi Reform (year 2003); this new law has increased short term employment opportunities after the degree, in fact, collaboration contracts are usually drawn up as soon as the degree has been awarded and are likely to be related to the degree. Considering the traditional indicator $I_t(S_2)$ we observe a similar pattern, even if, the permanent employment contracts show the maximum value of the indicator (0.81). As regard to the earning dimension $I_{m.f.}(\xi_2)_w$, the figures are inverted with respect to the $I_{m.f.}(\xi_1)_w$ indicator: the greater value of the indicator is related to collaboration contract. It seems to be reasonable that graduates are led to accept a collaboration contract, probably, coherently with the education path, also accepting an inadequate earning. These opposite results between the above indicators suggest a kind of trade off between the "theoretical perspective" of the quoted Biagi Reform and the real earning opportunities. For other types of contract (permanent and temporary employment) the mismatch indicator $I_{m.f.}(\xi_2)_w$ is more or less constant and it is significantly lower than the indicator of the collaborators.

If we consider the firm size as disaggregation variable, we can immediately observe that the size of two groups is actually very small (one worker and 50-99 workers). Considering the other classes of firm size and the multidimensional indicator $I_{m.f.}(\xi_1)_w$, it can be observed that the range of the value is quite narrow: the minimum value of the indicator is 0.35 and it is related to individuals employed in firms having 2-5 workers, the maximum value is 0.42 for individuals employed in firms having 15-49 workers. Observing these figures, we can conclude that firm size does not seem to have a relevant influence on the ξ_1 dimension. Quite different is the situation if we consider the traditional indicator; in fact, the values of the index are really very high for medium and large firms (the indicator is greater than 0.75, and it means that more than 75% have declared that the degree achieved was not requested for the employment); the indicator is otherwise equal to 0.62 for employees in small firms (2-5 workers); even if the value is relevant in the absolute sense, it is significantly lower than in other groups. Values related to the earning dimension, highlight two classes of workers: individuals employed in small (2-5 workers) and medium-large firms (15-49 workers) present a high value of the indicator (0.63 and 0.64); individuals employed in medium-small firms (6-14 workers) and large firms (100 or more workers) present significantly lower values of the indicator (respectively 0.49 and 0.45).

Another aspect to be considered concerns whether or not the graduates have changed job after graduation; we have defined three classes: individuals having the same job, individuals that after the degree have changed jobs, individuals that have found the first job after graduation. According to both the traditional and the multidimensional factors, the lower value of the indicators is, reasonably, related to individuals that had the first job after the degree. This effect can be explained by the fact that the achieved degree offers more significant opportunities to the graduates in the labor market. Moreover, the graduates are interviewed one year after the degree and it should be interesting to observe this effect in the long term.

With respect to the field of education followed, some interesting evidences can be observed: creating a ranking of the faculty according to the mismatch (either for the $I_{m.f.}(\xi_1)_w$ and $I_t(S_2)$), the position of the single faculty in the ranking is more or less the same; even if the traditional indicator is always higher than the multidimensional-fuzzy one. As expected, graduates in Law and Humanities present the highest values for the $I_{m.f.}(\xi_1)_w$ indicator (respectively 0.49 and 0.46) and quite high values for the $I_{m.f.}(\xi_2)_w$ indicator (respectively 0.48 and 0.68). On the other hand, it is reasonable that graduates in Pharmacy and Medicine present the lowest indicator for both $I_{m.f.}(\xi_1)_w$ and $I_{m.f.}(\xi_2)_w$. Economics and Political Science graduates are in the middle of the ranking for both the indicators, even if the mismatch is higher for the earning indicator (respectively 0.48 and 0.45) than for the $I_{m.f.}(\xi_1)_w$ indicator (respectively 0.36 and 0.37).

Comparing the indicators when regarding the group of graduates before and after the reform: differences are really negligible for both the multidimensional fuzzy indicators. This result implies that the university reform has a limited effect in solving the overeducation problems. As expected, the difference is more marked when observing the traditional indicator as the dichotomous sharing does not take into account the membership degree to the set of overeducated workers.

Let us now consider the group of the 298 graduates in medicine after the reform (the group labeled as paramedics). They present mismatch indicators $I_{m.f.}(\xi_1)_w$ and $I_{t.}(S_2)$ lower than the so-called total sample. In fact, it is quite reasonable that their job is related to the achieved degree; however, for them the mismatch multidimensional indicator on earning $I_{m.f.}(\xi_2)_w$, is remarkably greater than the one computed for the total sample. The reason could be the following: they are still working at the same level as before the degree, but now they retain that the level and the earning is no more adequate to their education level. Moreover, the females in this group are 74%, therefore the high degree of mismatch could be influenced by gender discrimination in earnings.

6. Some final remarks

In this paper, we start from the assumption that educational mismatch is a multidimensional and fuzzy concept, this is because we believe that educational mismatch is a very complex issue influenced by many factors, moreover different degree of membership of overeducation may exist that cannot be estimated by a dichotomous indicator. Taking into account the previous hypothesis, our main findings can be summarized as follows.

The values of the traditional indicator are systematically higher than the ones of the multidimensional-fuzzy indicator $I_{m.f.}(\xi_1)_w$. A dichotomous indicator does not capture the fuzzy aspects of the mismatch concept, in fact even if a worker is partially overeducated he is classified as totally mismatched while the membership function takes account of the mismatch degree. On the other hand, the multidimensionality of the new index measures the phenomenon on average, but still considering the specific the effect of each item. Moreover, the multidimensional-fuzzy indicator regarding the earning factor $I_{m.f.}(\xi_2)_w$ introduces new evidence for understanding educational mismatch. This is not possible with the traditional index.

In spite of the limits of the current dataset (our sample covers only graduates from the University of Siena) we feel that very interesting results have been discovered; such results can help to define and to measure overeducation among graduates.

In terms of policy implications, our findings make an important contribution to the discussion. It is well known, that the university reform in the Italian higher education system has introduced more levels in terms of university qualifications compared to the old university course which only had one level. The innovation intends to increase the productivity of the system, reduce the average length of studies and differentiate postsecondary tracks in relation to the labor market and the new professions. In our paper, the first graduates with a first–level university qualification have been observed and their outcomes in the labor market have been explored in terms of overeducation. Our findings show that the reform has not improved the educational match. In addition, the only difference is discovered in the traditional indicator and graduates with the new degree courses are at a disadvantage. This evidence suggests that further studies devoted to the analysis of the performance of Italian university students in the reformed system could be useful in order to validate our results.

Some other variables are also discovered to have an influence on overeducation. Women are more mismatched than men and the graduates in Pharmacy, Medicine and Engineering are the least overeducated in terms of the ξ_1 dimension, even if in terms of the earnings dimension (ξ_2) they have similar mismatches to the other fields of education. Self-employment and collaboration contracts reduce overeducation when the ξ_1 dimension is considered; on the other hand, when ξ_2 dimension is taken into account graduates with collaboration contracts are the most mismatched out of those having a job.

However, further developments of the study are necessary from a substantial point of view in order to measure the joint influence that the structural variables have on the educational mismatch. Therefore, future research could be direct to the specification of an econometric model using the proposed indicators as dependent variables. Then, the inferences from the model can be used for policy-makers decisions.

Finally, it could also be interesting to use the proposed multidimensional and fuzzy indicators, to analyze and to evaluate the impact that the recent reform, affecting most areas of the labor law (the Biagi reform of year 2003, based on the guidelines in the Europeans Employment Strategy), had in terms of overeducation. This implies a cohort study of graduates over more than one year that could be a very interesting topic for future research.

References

Aassve A., Betti G., Mazzuco S., Mencarini L. (forthcoming), Marital Disruption and Economic Well-being; A Comparative Analysis, *Journal of the Royal Statistical Society, Series A*.

Alba-Ramirez A. (1993), Mismatch in the Spanish labor Market: Overeducation?, *The Journal of Human Resources*, **28(2)**, pp. 259-278.

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

AlmaLaurea (2006), Condizione Occupazionale dei laureati. Indagine 2005.

Battu H., Belfield C.R., Sloane P.J. (1999), Overeducation among graduates: a cohort view, *Education Economics*, **7**, pp. 21-38.

Belfield C., Harris R. (2002), How well do theories of job matching explain variation in job satisfaction across education levels? Evidence for U.K. graduates. *Applied Economics*, **34**, pp. 535-548.

Bentler P.M., Bonnet D.G. (1980), Significance tests and goodness of fit in the analysis of covariance structures, *Psychological Bulletin*, **88**, pp. 588-606.

Betti G., Cheli B., Lemmi A., Verma V. (2006), Multidimensional and Longitudinal Poverty: an Integrated Fuzzy Approach, in Lemmi A., Betti G. (eds.), *Fuzzy Set Approach to Multidimensional Poverty Measurement*. New York: Springer, pp. 111-137.

Betti G., Verma V. (1999), Measuring the degree of poverty in a dynamic and comparative context: a multi-dimensional approach using fuzzy set theory, *Proceedings of the ICCS-VI*, Vol. 11, pp. 289-301, Lahore, Pakistan, August 27-31, 1999.

Blau F.D., Kahn L.M. (2000), Gender Differences in Pay, *Journal of Economics Perspectives*, **14**, pp. 75-99.

Bollen K.A. (1989), *Structural Equation with Latent variables*. New York: Wiley.

Browne M.W., Cudeck R. (1993), Alternative ways of assessing model fit, in Bollen K.A., Long J.S. (eds.), *Testing Structural Equation Models*. Thousand Oaks, CA: Sage, pp.136-162.

Cabral Veieira J.A. (2005), Skill mismatches and job satisfaction, *Economics Letters*, **89**, pp.39-47.

Cammelli A. (2007), Survey of Graduate Employment, Presented to the Conference on University-to-Work transition in Italy and in Europe, Bologna, 2-3 March 2007.

Cerioli A., Zani S. (1990), A Fuzzy Approach to the Measurement of Poverty, in Dagum C., Zenga M. (eds.), *Income and Wealth Distribution, Inequality and Poverty, Studies in Contemporary Economics*. Berlin: Springer Verlag, pp. 272-284.

Cohn E. (1994), The impact of surplus schooling on earnings: comment. *Journal of Human Resources*, **27**, pp. 679-682.

Di Pietro G., Cutillo A. (2006), University Quality and Labour Market Outcomes in Italy, *Labour*, **20**, pp. 37-62.

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

Dolton P., Vignoles A. (2000), The incidence and effects of overeducation in the U.K. graduate labour market, *Economics of Education Review*, **19**, pp.179-198.

Dubois D., Prade H. (1980), *Fuzzy Sets and Systems*, Academic Press, Boston, New York, London.

Filippone B., Cheli B., D'Agostino A. (2001), Addressing the Interpretation and the Aggregation Problems in Totally Fuzzy and Relative Poverty Measures, ISER Working Papers, n. 2001-22. Institute for Social and Economic Research, University of Essex, Colchester, UK.

Groot W. (1996), The incidence of, and returns to overeducation in the UK, *Applied Economics*, **28**, pp. 1345-1350.

Groot W., van den Drink H.M. (2000), Overeducation in the labor market: a meta-analysis, *Economics of Education Review*, **19**, pp. 149-158.

Hakstian A.R., Rogers W.D., Cattel R.B. (1982), The behavior of numbers factors rules with simulated data, *Multivariate Behavioral Research*, **17**, pp. 193-219.

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

Jarrell S.B., Stanley T.D. (2004), Declining bias and Gender Wage Discrimination? A Meta-regression Analysis, *Journal of Human Resources*, **39**, pp. 828-838.

Joreskog K.G., Sorbom D. (1993), *LISREL 8 user's reference guide*. Chicago: Scientific Software.

Kaiser H.F. (1960), The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, **20**, pp. 141-151.

Lucifora C., Meurs D. (2006), The Public sector pay gap in France, great Britain and Italy, *Review of Income and Wealth*, **52**, pp. 42-59.

Nasselroade J.R., Baltes P.B. (1984), From traditional factor analysis to structural-causal modelling in development research, in Sarris V., Parducci A. (eds.), *Perspective in psychological experimentation: toward the year 2000*. Hillsdale, New Jersey: Lawrence Erlbaum, pp. 267-287.

Raghunathan T. E., Lepkowski J., Van Voewyk J., Solenberger P. (2001), A Multivariate Technique for Imputing Missing Values Using a Sequence of Regression Models, *Survey Methodology*, **27**, pp. 85-95.

Rubb S. (2003), Post-college Schooling, Overeduction, and Hourly Earnings in the United States, *Education Economics*, **11**, pp. 53-72.

Whelan C.T., Layte R., Maitre B., Nolan B. (2001), Income, deprivation and economic strain: an analysis of the European Community Household Panel, *European Sociological Review*, **17**, pp. 357-372.

Zadeh L.A. (1965), Fuzzy Sets, *Information and Control*, **8**, pp. 338-353.