

HOW TO REWARD TRAINING AGENCIES ACCORDING TO THEIR PERFORMANCE? A PROPOSAL FOR ITALIAN REGIONS

Elena Cappellini*
Silvia Duranti*
Valentina Patacchini*
Carla Rampichini**
Nicola Sciclone*

* *Irpet* ** *University of Florence*

Introduction

In the last years, the trend towards activation has been one of the major issues in welfare and labour market reforms in Europe. Italy has lacked for a long time a strong net of activation policies for the unemployed, but the latest reforms (Fornero law in 2012, Jobs Act in 2014) have placed great emphasis on the need to invest in Public Employment Services to make labour market more inclusive. At the same time the European Union, through the European Social Fund, has made available the necessary financial resources to promote active policies, especially those related to training, considering them crucial for the development of human resources in a knowledge-based economy.

The increasing focus on activation policies and in particular on training raises the issue of better evaluation, not only of training programs themselves, but also of their providers. Being that of training a natural quasi-market, evaluation of providers¹ is the key factor for shaping an efficient governance structure for activation and training services, which is a major determinant of its success (Oecd, 2005).

In Italy, still little attention is given to the analysis of different providers, although some empirical works on the Regional employment services of Lombardy exists (Gori, 2000; Ballarino et al., 2001), as well as some others in the field of school effectiveness (Cipollone et. al., 2010; Bratti and Checchi, 2013; De Simone and Gavosto, 2013; Conti et al., 2016).

This paper proposes for the Italian context a governance framework, which relies on training providers' evaluation to regulate the market of agencies and efficiently allocate resources between them.

The rest of the paper is organised as follows. Section 1 will describe the Italian institutional framework in the field of vocational training, both at the national and regional level. In section 2 a new governance framework for the Italian regional training system is designed while the subsequent sections describe in detail the different steps to implement it. In particular, Section 3 describe how to create a system of profiling, Section 4 illustrates how to rank providers according to their adjusted performance and Section 6 simulates the allocation of resources under the new system.

¹ It is important to stress that we speak here of relative evaluation. The objective is not to evaluate the absolute effectiveness of training providers, but the relative performance of each provider compared to the others.

1. The Institutional framework

In Italy training is competence of the Regions, which, within certain standards defined at the national level², have defined models characterised by different levels of planning centralisation and training providers' regulation. However, the general context is one of quasi-market, where governments tender for training providers, which can contend public resources available to run courses.

The current Tuscan model is based on a bottom up planning of training supply, since the content of tenders is mainly based on the needs expressed by social parties and gathered at the Regional or Provincial level. Calls for tenders are open to all agencies having obtained accreditation and financial resources are allocated between applicants according to criteria related to the quality of the project.

The current model is based on accreditation rules where the performance of agencies plays no role. In particular, necessary and sufficient conditions to obtain accreditation is the possession of a minimum requirement, based on four criteria: organizational and administrative structure, logistics, relations, performance. The latter however, does not consider any indicator of job placement, being made up of 3 indicators concerning: rate of drop out, rate of trained, efficiency of spending. Once obtained accreditation, penalties affect training providers only in the event of formal errors and deficiencies and the revocation or suspension of accreditation only occurs when minimum requirements relating to formal criteria are not respected; any real role is attributed to indicators related to effectiveness.

A different model is the one implemented in Lombardy, especially in what concerns regulation. Indeed, the role of demand in shaping supply is maximum, because a system of vouchers (Dote), guarantees to individuals freedom of choice between the alternatives autonomously proposed by training providers, which therefore have incentives to adapt to demand. In this context regulation is extremely relevant. For this reason, the regulation of providers is based on a system of Rating, which goes beyond accreditation based on minimum standards towards competition and improvement in quality. In particular, the rating model is configured as a further development of the accreditation system and aims to measure the relative positioning of each training provider with respect to employment performance of trainees. Therefore, the model is not aimed at identifying agencies to be accredited or accredited providers that do not meet the required conditions: under investigation in the rating are the only agencies that have achieved accreditation to provide training services.

The model of Lombardy is a good benchmark for the improvement of the effectiveness of the Tuscan training system. However, the latter would be less market-oriented and based on a top-down planning of training supply, developed through calls for tenders for training courses. The idea is to establish a central governance of the content of supply (which courses to finance and how), leaving to the market the provision. In order to provide incentives for an effective training system, the current model of selection of providers is a good starting point, but it requires upgrading to make an assessment that goes beyond mere respect of formal criteria. Indeed, it is necessary to associate to formal selection a more substantial rating mechanism, which, as in the experience of Lombardy, gives more importance to the performance of providers. Being one of the

² For example, the basic principles of accreditation, inspired to the European framework of reference for quality assurance of education and training (EQARF), are defined at the national level in order to provide some indicators that guarantee on the quality of providers of education and training. In particular, the Italian system is based on five criteria that provide specific instructions to regional authorities, competent in training issues, to define the specific use of accreditation:

- general management of the institution;
- financial situation of the institution;
- staff characteristics for teaching, training and administration;
- effectiveness and efficiency of previous training activities;
- links and contacts (to schools, employers, employment services, municipal authorities, NGOs, etc.) at local level.

Regional governments can add on or expand criteria for accreditation and are called to evaluate whether training providers fulfil these requirements.

primary goals of training (in particular training for the unemployed) the increase in employability, the evaluation of performance of a provider should be based on its capacity to train people and get them back to work. As in the Lombardy experience, the rating model is configured as a further development of the accreditation system (which would continue to exist selecting agencies on the basis of formal or financial criteria), aimed at promoting an allocation of resources which reflect performance and stimulates competition between providers and thus an overall improvement of the system³.

The proposed model could be applied to any Regional system of Italy, although it is simulated in the next sessions for the context of Tuscany.

2. A proposal for a new governance framework of the training system

The management of the training system first requires the establishment of an adequate model of selection of training providers, able to select the most successful ones in order to effectively distribute resources.

The accreditation process is a good starting point in this process, but for the sake of the system's effectiveness and efficiency, this first selection needs to be complemented by a more substantial system of rating, based on the results of funded activities. The rating model is a development of the accreditation system and aims to measure the relative positioning of each training provider in a performance ranking net of external factors. Therefore, the model is not aimed at identifying sites to be accredited or accredited providers that do not meet the required conditions: only the providers that have already achieved accreditation are under investigation.

The rating can be used to select the best performers, to whom a certain amount of resources can be reserved. Since new entrants in the training system cannot be evaluated based on past performances, it is necessary to leave part of total resources to be contended by all providers.

The following is an example of how to allocate the total available resources: 40% to all operators (including outsiders) irrespective of previous re-employment performance, 40% to insiders having had a high or medium performance in the last year and 20% to best performers. Calls for projects would thus be differentiated by segment and, among each segment, financial resources would be still allocated between applicants according to criteria related to the quality of the project.

Differences between the current and the proposed model are clearly identifiable in Fig. 1, where major changes are depicted in light blue boxes.

³ Lam (2007) reports for Australia an improvement of 15 percentage points in the average outcome rate of Employment Services after only 18 months of implementation of the "Star Rating System".

Fig. 1 Current vs. Proposed system of training

Current model	Proposed model
<p>Accreditation:</p> <ul style="list-style-type: none"> • formal criteria not concerning employment performance • penalties in case of formal errors and administrative deficiencies 	<p>Formal accreditation</p>
<p>Call for projects</p>	<p>Employment performance evaluation: allocation of agencies into 3 segments (best, medium and worst performers)</p>
<p>Projects proposals from training agencies</p>	<p>Calls for projects by segment:</p> <ul style="list-style-type: none"> • 40% contestable by all agencies • 40% contestable by not worst agencies • 20% contestable by best performing agencies
<p>Projects proposals from training agencies</p>	<p>Projects proposals from training agencies</p>
<p>Projects evaluation, selection and financing by the Regional administration</p>	<p>Projects evaluation, selection and financing by the Regional administration</p>

In summary, a model of governance based on the selectivity and transparency in allocating resources between providers should be based on three steps.

First, the creation of a profiling system that summarizes the level of disadvantage in the labour market of each trainee. Such a tool is necessary to correctly evaluate net performance of training operators and to prevent creaming or cherry peaking phenomena; a proposal for this is described in section 3, where a statistical approach is used.

Second, the creation of a rating system of providers based on their net performance in terms of employment outcomes; an example of how this can be made is provided in section 4, where an econometric approach has been adopted to obtain a ranking of training providers.

Third, the allocation of resources based on the reward of best performers in terms of employment outcomes. A simulation of this procedure for 2013 is proposed in section 5, where the new hypothetical distribution of resources is compared to the actual one. In the simulation, real shares of financing in 2013 among segments are maintained, assuming that the ability of providers to win tenders based on the quality of projects would remain the same in the new model.

3. Profiling system

To summarize the characteristics of training users we have created a tool for profiling users based on their distance from the labour market. Indeed, the methodology of profiling is based on the idea that employability is predictable given unemployed characteristics and, in particular, the length of unemployment. The goal is to

identify the characteristics that best define the profiles of people furthest away from the labour market and therefore harder to be reallocated and summarize them in a single indicator (O'Connell et al. 2009; Rudolph and Konle-Seidl, 2005; Lam 2007; Hasluck, 2004).

The methodology used is rather similar to the one adopted in the Italian region of Lombardy and described in Arifl (2013). The key variables for the creation of profiling scores are: sex, age class, education level and type of unemployment. These variables are used as covariates in a logit model for the estimation of the probability to find a job for an unemployed person within 12 months since arrival at the Public Employment Service⁴. Technically, we estimate a saturated model, namely a model in which there are as many estimated parameters as type of subjects, as defined by the interaction of all covariates⁵. By definition, this model will lead to a perfect fit, and it allows predicting the employment probability for all the combinations of personal characteristics. The resulting employment probability of each profile is used as profiling score, representing the distance from the labour market and ranging between 0 (maximum level of disadvantage) and 1 (minimum level of disadvantage), as shown in Table 1.

Table 1.
Profiling scores

Age class	Sex	Education level	First-time job-seeker	Short term unemployed (12 months or less)	Long term unemployed (more than 12 months)
Under 29	M	Compulsory education	0,46	0,59	0,44
		Higher education	0,46	0,75	0,55
	F	Compulsory education	0,39	0,52	0,37
		Higher education	0,46	0,74	0,52
30-49	M	Compulsory education	0,29	0,56	0,43
		Higher education	0,26	0,69	0,49
	F	Compulsory education	0,24	0,54	0,41
		Higher education	0,21	0,68	0,42
50+	M	Compulsory education	0,16	0,38	0,26
		Higher education	0,13	0,51	0,37
	F	Compulsory education	0,18	0,44	0,33
		Higher education	0,07	0,54	0,23

4. A system of rating of training providers

4.1 Sources of data

⁴ The population on which the logit regression has been estimated is represented by all people signed up to unemployment list by Public Employment Services in the period 2008-2012.

⁵ The importance of estimating such a model instead of a linear model can be explained by a simple example: in the younger age class being a non-qualified woman is a disadvantage, while later it becomes an advantage, guaranteeing better employment chances. In a linear regression this heterogeneity would not show up, and being a non qualified woman would probably look like a disadvantage.

The primary source of data for evaluating the training system is the database of the European Social Fund, which represents the main source of funding for the regional training system. From this dataset we have extracted the training courses financed by Tuscany in 2013, selecting only those for the unemployed⁶. We thus consider a population of 181 providers, 761 courses and 8,441 unemployed trainees. The dataset contains information on the trainees, specifically concerning sex, age, nationality, education level, previous work experience and duration of unemployment. Information on courses are also very rich and concern duration, thematic content, class size, cost to public finance, data of beginning and end. Moreover, we use the Compulsory Communications System of administrative data on employment dynamics, which record all the activations, transformations, fixed-term extensions and anticipated terminations of employment relationships between any worker and employer since the beginning of 2008.

Merging labour market administrative data with the dataset on trainees, courses and providers it is possible to check the employment outcomes of all unemployed people having attended a training activity. A limitation of the joined data set is the lack of information on self-employment: placement rates are therefore net of activation as self-employed. The choice to use only administrative data comes from the need of developing standard and replicable procedures in a new framework of governance. Another advantage of administrative data instead of ad hoc surveys for the analyses of employment outcomes is the accuracy of information: indeed, the Compulsory Communications System of administrative data on employment dynamics contains information on all jobs following the end of the course, their timing and length and many other characteristics (sector of activity, contract and qualification).

4.2 Theoretical framework and methodology

There are several options to rank based on their performance. Among the exiting options at the international level⁷, the Australian methodology is the nearest to our proposal (Struyven and Steurs, 2005; Lam, 2007; Australian Government-Disability Employment Services, 2013), since it uses regression-based estimates of performance to rank providers. The aim is to enable fair comparisons between providers, by taking into account of factors outside their control. Among these factors, there is certainly the type of user. For this reason, such a governance model needs to rely on a profiling system as the one described above.

To rank training providers, the approach of the school-effectiveness research (SER) was used. In particular, the goal is to rank training providers according to their Type B effectiveness (Raudenbush and Willms (1995), which is the actual performance of providers, net of factors that are exogenous and thus out of their control. These studies usually make use of multilevel methodologies to take into account the hierarchical nature of data and to exploit second level residuals to measure school effectiveness (Raudenbush and Willms, 1995; Goldstein, 1997; Grilli and Rampichini, 2009). Indeed, second level residuals are considered as the “net effects” of training providers. This approach is not new for the evaluation of effectiveness of courses and training providers in the Italian context. The same approach has been adopted by Gori (2000) for the Lombardy Region, by Gori et al. (1993) and Regione Toscana, Giunta Regionale, Osservatorio Regionale del mercato del lavoro (1997) for Tuscany, and, more recently, by Paccagnella (2006) for the Autonomous Province of Bolzano.

The measurement of the training providers' performance first requires the identification of one or more measurable dimensions for which the ranking has to be produced (placement rates, time to find the first job, days worked in the year following the course). International and regional experiences offer several examples of different outcome measures of effectiveness upon which to evaluate providers (Lam, 2007; Finn, 2009). In this paper, we consider as outcome variable the placement rate in the 12 months following the end of the course.

⁶ Moreover, the database has been cleaned from courses specifically addressed to disabled, identified by those having a percentage of disabled higher than 50% of total trainees. For this courses employability is only one of the goals, having as primary aims social inclusion and acquisition of basic skills.

⁷ See Ocse, 2005 for a more general review on Public Employment Service evaluation.

The methodology used consists to fit a two-level random intercept logit model (Goldstein, 2003) to properly take into account the hierarchical structure of the data, i.e. trainees nested into training providers.

Let Y_{ij} be the binary response, i.e. $Y_{ij} = 1$ if the i -th subject of the j -th agency find a job and zero otherwise, where $i = 1, \dots, n_j$ denotes the number of trainees (level 1 units) nested within the second level unit (or cluster) j , i.e. the training provider, $j = 1, \dots, J$, and J is the total number of considered agencies. Given the success probability $\pi_{ij} = P(Y_{ij} = 1 | \mathbf{x}_{ij}, u_j)$, the model is specified as follows:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \boldsymbol{\beta}\mathbf{x}_{ij} + u_j$$

where \mathbf{x}_{ij} is the vector of covariates for the i -th subject of the j -th cluster (including a constant term for the intercept) and $\boldsymbol{\beta}$ is the vector of fixed parameters (including the intercept). The residuals u_j are assumed to be independent and identically distributed across clusters with a normal distribution and common variance σ_u^2 . Model fitting is performed using the melogit procedure of Stata (StataCorp, 2015).

The model allows considering covariates at both levels. At the trainee level, we considered individual characteristics and job status, as reported in Table 4, while as second level explanatory variables we included only those factors assumed to affect the outcome but outside of the control of the agency. In particular, we considered aggregated characteristics of trainees and labour market conditions, measured by the year and the area where the course has taken place. The characteristics of trainees are included both at the individual and at the agency level.

The characteristics of the course in terms of content (vocational versus non-vocational) and length have been included in the model. This is consistent with a top-down governance framework, where the Region plans the allocation of funding between types of training (for young people or mature ones; vocational or general, structured, light, or intensive), leaving to providers only the detailed planning of single courses. For this reason, providers cannot be penalised for having organized light courses, which notoriously guarantee lower employment outcomes, but which have been planned and financed by the regional government. On the contrary, a more market-oriented system, where the planning and supply of training is fully in charge of providers (which only have the goal of employability, whatever this goal is achieved), would require a rating system where performance is considered gross of course characteristics⁸.

Several characteristics of the providers could be considered in the analysis: type of provider (school, firm, local governments, no-profit organization), size in terms of income or employees, degree of courses specialisation, employees, frequency of funding with public resources⁹. However, for the purpose of this analysis they do not need to be controlled for when estimating a net performance: each training provider has as primary goal the increase in the employability of trainees, irrespective of how this goal is pursued.

Our estimation procedure follows four steps: in the first step, we estimate an "empty" model, to decompose the total variance into the trainee-level (within) variance and training provider-level (between) variance and test the relevance of the latter, while in the second, third and fourth steps we add explanatory variables respectively at individual, course and provider level.

Table 3 reports the estimated coefficients and the probabilities computed for the reference subject (all covariates equal to zero), for varying values of the level two-residuals. The reported likelihood-ratio test compares the random intercept model to ordinary logistic regression (Berkhof and Snijders, 2001) and is highly significant for these data, confirming an "agency effect", which can change the employment probability of a trainee from 48% to 75% according to the gross training provider effectiveness (model without covariates).

⁸ It is the case of Australia, but also of Lombardy.

A similar model has been tested with our data and results differ significantly from the basic model. In particular, more than 20% of providers resulting best performers in the basic model, leave the top of the ranking in the alternative one.

⁹ Although the dataset contains only the name and the VAT identification number of training providers, the original data set has been enriched with the information contained in Iripet's database on firms: employees, income, type of provider (firm, no profit).

Once controlled for exogenous variables, providers still appear to perform significantly differently in regards to rate of re-employment. Indeed, in the full model (Column D of Table 3) the likelihood-ratio test is still highly significant, pointing out unexplained second level variability. Consistently, the 95% coverage interval of the predicted employment probability for the reference trainee ranges from 9% to 33%, according to the training provider net effectiveness (net of users', courses' and labour market characteristics). To give an idea of the effect of unobserved factors at the training provider level Table 4 reports the predicted probability of employment considering three types of providers (an average effective provider, a low effective provider and a highly effective one) and four typical profiles of trainee.

Table 3.
Results of the multilevel logistic regression

	Empty	Individual	Individual+Course+Contest	Individual+Course+Contest+Agency
Covariates	A	B	C	D
Disabled		-0,406	-0,362	-0,312
Profiling score		1,91***	1,85***	1,77***
Vocational course			-0,00915	-0,00066
Diluted course			-0,0619	-0,0408
Intensive course			0,036	0,0372
Light course			-0,191*	-0,182*
Structured course			0,26***	0,23***
Employment rate in Local Labour System			-0,0749***	-0,0708***
Average profiling score in the agency				2,15**
Percentage of disabled in the agency				-0,0187
Constant		-1,05***	-0,479**	-1,53***
Probability of employment for the reference subject:				
in an average agency (u=0)	47%	26%	38%	18%
in a low performing agency (u=2sd)	25%	13%	21%	9%
in a high performing agency (u=-2sd)	70%	45%	60%	33%
LR test vs. logistic regression:				
chibar2(01)	1075,97	668,6	475,74	334,4
Prob>=chibar2	0	0	0	0
<i>ICC</i>	<i>6,7%</i>	<i>5,4%</i>	<i>5,5%</i>	<i>4,8%</i>

Table 4.

Predicted re-employment rates for typical profiles of trainees in differently effective training providers. Full model (column D of Table 4)

	Average agency (sd=0)	Low performing agency ($u=-2*sd$)	High performing agency ($u=2*sd$)	Difference
Woman over 50, long term unemployd, with only compulsory education	41%	23%	61%	38%
Young man with discontinuos work experience, short term unemployd	80%	64%	90%	26%
35 years-old woman in reactivation, with a degree and long term unemployd	50%	30%	69%	39%
Young man early shool leaver looking for his first job	57%	37%	75%	38%

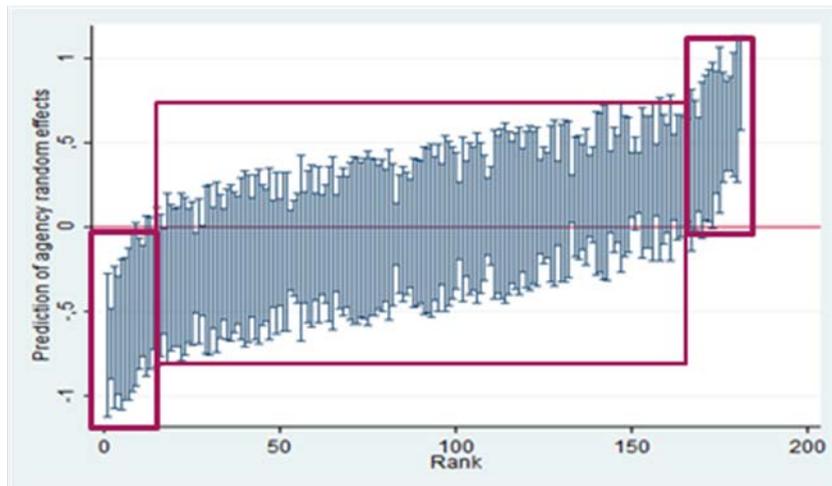
Note. Predicted probabilities with observable course and agency characteristics equal to zero, i.e. the base categories of level 2 covariates: non vocational course, with average duration in terms of length and hours, and a training providers with no disabled and average minimum profiling score.

4.3 Ranking of training providers

We ranked Tuscan publicly funded training providers according to their effectiveness, measured by Empirical Bayes (level 2) residuals of the full model (column D of Table 3). Figure 2 shows the predictions of level 2 residuals from the full model alongside with their comparative confidence intervals. The training providers in the lower left and upper right part of the graph have a predicted residual significantly different from zero. The latter, are those to be considered the best performers of the Tuscan training system, because they show placement rates significantly higher than those predicted by the model based on observable characteristics of users and courses. These 19 agencies are those to be rewarded by deserving them a quota of public resources.

However, who are the best performers? Descriptive statistics show that they are mainly firms (instead of schools, nonprofit organizations and local governments), whose core activity concerns training. This information is not trivial, since around half of training providers have not as core business training itself. Looking at public resources in the observation period, best performers do not seem to have received on average much more than other providers, and this is consistent with the lack of any kind of performance evaluation in the current model.

Fig. 2.
Ranking of training providers according to second level residuals from multilevel regression



5. Allocation of resources

To simulate the impact of such a governance framework of training providers on the allocation of resources between insiders, the ranking has been used to select best performers as those showing significantly higher placement rates than those predicted by the model.

The simulation of the new allocation of resources has been done relying on two strong assumptions.

The first assumption concerns the maintenance of the relative quotas of resources between operators within segments. Being the final allocation of resources among segments based on the quality of the projects, it is assumed that a provider's ability to propose high quality projects would remain unchanged in the new model.

The second assumption is the one of complete elasticity of supply: the more resources are available, the more agencies use them. This might be a strong assumption for best performers, which could almost double their amount of public resources.

Table 6 shows the simulated allocation of public resources applying the rating system to 2013. Only 40% of total resources (about 7,6 million euros) are contestable between all training providers (including new entrants); the 40% of resources (about 7,6million euros) are contestable only between those agencies having proved more performing than the average (agency residual higher than zero), while the remaining 20% (about 3,8 million euros) is reserved to best performers (agency residual significantly higher than zero). This means that each provider theoretically disposes of 42.021euros, which become 86.761 euros (44.740 euros more) for those providers not selected as worst performers and jump to 286.913 (200.152 euros more) for best performers.

Simulating this new allocation on effective agency data for the 2013, we found that changes in the distribution of resources are concentrated on best and worst performers, while the majority of training providers do not change significantly the amount of public resources received. In the new model, financial resources available to best performers totally increase by 97% compared to the effective distribution in 2013. On the contrary, financial resources available to worst performers totally decrease by more 60%.

Thus, the distribution of resources under the new model is far less homogeneous; as show in table 6, in the actual distribution half resources were held by only the 20% of training operators, a percentage that would be reduced to 15% in the new model.

Table 5
Allocation of public resources applying the rating system

	All	Not worst	Best
Total reserved amount	€ 7.605.794	€ 7.605.794	€ 3.802.897
Number of obs.	181	170	19
Individual average amount	€ 42.021	€ 44.740	€ 200.152
Individual cumulated amount	€ 42.021	€ 86.761	€ 286.913

Table 6
Distribution of public resources applying the rating system

<i>% resources</i>	<i>% training operators in the old model</i>	<i>% training operators in the proposed model</i>
20%	6%	3%
40%	14%	10%
50%	20%	15%
60%	28%	22%
80%	48%	42%
100%	100%	100%

Conclusions

The governance structure for activation and training services is a major determinant of success. For this reason, this paper proposes for the Italian context a governance framework inspired to the principles of regulation of providers and efficiency of resources allocation.

The idea is to establish a central governance of the content of supply (which courses to finance and how), leaving to the market the provision. In order to provide incentives for an effective training system, it is necessary to associate to formal selection a more substantial rating mechanism, which gives more importance to the performance of providers. Since the increase in employability is one of the primary goals of training for the unemployed, the evaluation of performance of a provider should be based on its capacity to train people and get them back to work. Therefore, the proposed methodology is based on a regulation of the market, which rewards agencies with better employment performances, in order to stimulate competition for an overall improvement of the system.

The proposed methodology, although based on clear and transparent criteria, leaves room for political choices.

First of all, a different outcome variable can be chosen. In this paper we chose a simple measure (at least a job entry in the 12 months following the course), but it is also possible to consider more specific, or qualitative measures. Possible future developments of this work include ratings based on alternative outcomes, which consider time (e.g. time to find a job or duration of the job found) and qualitative aspects; in particular the type of contract would be an interesting aspect to be considered, possibility through the use of a multinomial multilevel regression model.

Secondly, also the inclusion of covariates in the model leaves room for political choices, because it depends on the type of training "market" one wants to develop. In this paper, we assumed a centrally planned training supply, where, therefore, some characteristics of the course are exogenous to the agency, and thus need to be controlled for. On the contrary, a more market-oriented system, where the planning and supply of training is fully in charge of providers (which only have the goal of employability, whatever this goal is achieved), would require a rating system where performance is considered gross of course characteristics.

Finally, given the predicted agency's model-adjusted performance, the allocation of resources between training providers depends on how strict the identification of best performers is (only those with a performance significantly different from the average, or all those with a positive residual, or the first quartile in the ranking ecc.) and on the quota reserved to them.

The possibility to test different versions of the proposed model (along the aforementioned aspects) makes it flexible enough for the effective implementation by the policy maker.

References

Arifil (2013), *La Dote Unica Lavoro Il sistema delle fasce ad intensità di aiuto Descrizione della metodologia*, Newsletter Arifil N.2 - Ottobre 2013.

Australian Government-Disability Employment Services (2013), *DES Star Ratings Methodology Advice*. https://docs.employment.gov.au/system/files/doc/other/disability_employment_services_star_ratings_methodology_advice.pdf

Ballarino G., Checchi, D., Fiorio C., Iacus S., Leonardi M. e Porro G. (2011), "La valutazione dell'efficacia del "sistema delle doti" della Regione Lombardia: modelli statistici e criticità nella progettazione", *Rassegna Italiana di Valutazione*, n. 49.

Berkhof, J. and Snijders, T. A. B. (2001), Variance Component Testing in Multilevel Models, *Journal of Educational and Behavioral Statistics*, Vol. 26, No. 2 (Summer, 2001), pp. 133-152.

Bratti, M. and D. Checchi (2013). *Re-testing PISA Students One Year Later: On School Value Added Estimation Using OECD-PISA*. IZA Discussion Papers 7722, Institute for the Study of Labor.

Cipollone, P., Montanaro, P. and P. Sestito (2010). "Value-Added Measures in Italian High Schools: Problems and Findings". *Giornale degli Economisti*, 69(2), 81-114.

Conti E., Duranti S., Sciclone N, and Rampichini C. (2016). "The future has early roots: learning outcomes and school effectiveness in Tuscany's primary education system", in *Youth and the Crisis: unemployment, education and health* (Coppola, G. and O'Higgins, N.), Routledge, pp. 89-108

De Simone, G. and A. Gavosto (2013). *Patterns of Value-Added Creation in the Transition from Primary to Lower Secondary Education in Italy*. FGA Working Paper n. 47, Fondazione Giovanni Agnelli (Turin).

Finn, D. (2009), *Differential pricing in contracted out employment programmes: review of international evidence*. In: Research report . Department for Work and Pensions.

Goldstein, H. (1997) "Methods in School Effectiveness Research", *School Effectiveness & School Improvement*, Vol. 8, Issue 4, pp.369-395.

Goldstein, H. (2003), *Multilevel statistical models*, 3rd edn. Arnold, London.

Grilli, L. and Rampichini, C. (2009), "Multilevel models for the evaluation of educational institutions: a review", in Bini M., Monari P., Piccolo D., Salmaso L., *Statistical Methods for the Evaluation of Educational Services and Quality of Products*, pp. 61-80, Heidelberg: Physica-Verlag HD.

Hasluck, C. (2004), *Targeting Services in the Individual Customer Strategy: the role of profiling. A review of research evidence*, for Department for Work and Pensions. *Italiana di Valutazione*, Vol. XV, N. 49, pp. 39-61.

Lam, M. (2007), *Performance measures for welfare-to-work programmes: The relevance of Australian star ratings to the UK*, Work Directions UK, London.

Nuremberg, January 12-14, 2005 Supported by EU Commission (DG Empl) VP/2004/007.

O'Connell, P. J., MCGuinness, S., Kely, E., Walsh, J. (2009), National profiling of the unemployed in Ireland, *Research series number 10, July 2009*, The Economic and Social Research Institute, Dublin.

OECD (2005), "Labour Market Programmes and Activation strategies: Evaluating the Impacts", in OECD Employment Outlook, OECD Publishing.

Paccagnella, O. (2006), Comparing vocational training courses through a discrete-time multilevel hazard model, *Statistical Modelling*, 6: 119-139.

Raudenbush S. and Willms, J. (1995), "The estimation of school effects", *Journal of Educational and Behavioral Statistics*, Vol. 20, pp. 307-335.

Rudolph, H. and Konle-Seidl, R. (2005), *Profiling for Better Services*. Report on the European Profiling Seminar.

StataCorp (2013), *Multilevel Mixed-Effects Reference Manual*, Stata: Release 13. Statistical Software. College Station, TX: StataCorp LP.

Struyven, L. and Steurs, G. (2005), "Design and redesign of a quasi-market for the reintegration of jobseekers: empirical evidence from Australia and the Netherlands", *Journal of European Social Policy*, 15, 3, 211–29.

Wooldridge (2013), "A Binary Dependent Variable: The Linear Probability Model". *Introductory Econometrics: A Modern Approach*, 5th international ed.. Mason, OH: South-Western. pp. 238–243.