Individual Disadvantage and Training Policies: The Makings of "Model-based" Composite Indicators

Rosanna Cataldo, Maria Gabriella Grassia, Natale Carlo Lauro, Elena Ragazzi, Lisa Sella
Individual Disadvantage and Training Policies: The Makings of "Model-based" Composite Indicators

Rosanna Cataldo 1, Maria Gabriella Grassia 2, Natale Carlo Lauro 3, Elena Ragazzi 4, Lisa Sella 5*

* Corresponding author:
Mail: lisa.sella@ircres.cnr.it
Tel.: +39 0116824926

ABSTRACT: In evaluating a policy, it is fundamental to represent its multiple dimensions and the targets it affects. Indeed, the impact of a policy generally involves a combination of socio-economic aspects that are difficult to represent. In this study, regional training policies are addressed, which are aimed at recovering the huge gaps in employability and social inclusion of weak Italian trainees. Previous counterfactual estimates of the net impact of regional training policies show the mess to observe and take into account the manifold aspects of trainees’ weakness. In fact, the target population consists of very disadvantaged individuals, who experience hard situations in the labour market. To overcome this shortfall, the present paper proposes a Structural Equation Model, that considers the impact of trainees’ socio-economic conditions on the policy outcome itself. In particular, the ex ante human capital is estimated from educational, social and individual backgrounds. Then, labour and training policies augment the individual human capital, affecting labour market outcomes jointly with individual job search behaviour. All these phenomena are expressed by a wide set of manifest variables and synthesised by composite indicators calculated with Partial Least Squares SEM. The makings of SEM are appraised, applied to the case of trainees in compulsory education.

KEYWORDS: Impact evaluation, labour policies, composite indicators, structural equation models.

JEL CODES: C13; I24; J48

1 Università Federico II (Napoli), mail: rosanna.cataldo2@unina.it.
2 Università Federico II, Dipartimento di Scienze Sociali, Vico Monte della Pietà, Napoli, mail: mgrassia@unina.it.
3 Università Federico II, Dipartimento di Matematica e Statistica, via Cintia 26, Napoli, mail: clauro@unina.it.
4 National Research Council of Italy, Research Institute on Sustainable Economic Growth, CNR-IRCrES Via Real Collegio 30, Moncalieri (TO), mail: elena.ragazzi@ircres.cnr.it.
5 National Research Council of Italy, Research Institute on Sustainable Economic Growth, CNR-IRCrES Via Real Collegio 30, Moncalieri (TO).
CONTENTS

1. INTRODUCTION AND MOTIVATION .................................................................................................................. 5
2. SEM METHODOLOGY AND THE MIXED TWO-STEP APPROACH .............................................................. 6
3. A SEM APPROACH TO EVALUATE INITIAL TRAINING .................................................................................. 8
   3.1 The model ....................................................................................................................................................... 9
   3.2 Survey data .................................................................................................................................................. 10
   3.3 Estimation results ........................................................................................................................................ 11
4. CONCLUSIONS ................................................................................................................................................. 12
5. REFERENCES .................................................................................................................................................... 14
1. INTRODUCTION AND MOTIVATION

The work based on structural equation models we present in this paper assesses the entry and exit conditions of individuals enjoying a training course. This statistical exercise has to be put in the perspective of our wider work, which is the evaluation service for the POR-FSE 2007-2013. In this context we were charged to run a yearly placement analysis, aimed at estimating the net impact in terms of employment of training policies, based on a quasi-experimental design.

The European Social Fund (ESF, FSE in Italian) promotes a wide range of social and labour policies aimed at enhancing social inclusion in Member States, but we were asked to concentrate on a sub-set of training policies sharing some common characteristics. All courses considered in the evaluation exercise, were long lasting (ranging from some 300h modules up to 3 year courses) full time courses, they provided a final certificate (either professional qualification or specialisation), they included a compulsory traineeship period and they were mostly addressed to unemployed people. For the sake of generality, courses for highly disadvantaged groups (e.g. prison inmates or disabled individuals) were not included in this study. Three main types of training were present: qualification courses for young people, basic qualification courses addressed to migrants and low-educated adults, and advanced specialisation courses addressed to high-educated adults with qualifications not appreciated by employers. On the whole, they are mostly weak targets, who experience difficult school experiences or labour market transitions and turn to vocational training to find new job opportunities.

In this paper we concentrate on training courses for young people eligible for compulsory education (CE, Sella Ragazzi 2011). In Italy the school age ends at 16, but young people are entitled to receive further training (even if they access the labour market) until 18. CE courses join classes of general disciplines, such as language and mathematics, to professional classes and laboratory traineeship. This is why they are generally attended either by students coming from low income families, who appreciate the shorter duration respect classical education and the professional finalisation intended to ease an early labour market entrance, or by students suffering of learning disabilities or of previous school failures, who enjoy the inductive and practical pedagogical approach (Ragazzi 2010) and tailored pathways (Ragazzi 2007, Lauro Ragazzi 2011). In this sense CE training has to be intended as a mixed policy, aiming at enhancing both human capital and employability.

In our evaluation activity we want to assess the net impact of training policies wherever possible, but we cannot rely on a random trial design (random assignation to treatment or control group). So we were forced to adopt a quasi-experimental approach, designing a control group made of individuals which are very similar to the trainees. This is very difficult in a situation where the courses tend to be attended by weak individuals, which cannot be compared to an average individual. Our choice (Falavigna et al. 2015) was to rely
on drop-outs and no shows (Bell 1995) as a way to control for differences in unobservables, and indeed Heckman (1999) tests excluded the existence of significant selection bias (Ragazzi 2014). But even so the theme deserves wider attention; there is strong suspect that the treated have different characteristics because the variables describing the disadvantage are often not observable and measurable and because it is a multidimensional phenomenon.

Moreover, in some cases such as the object of our paper, it is impossible to design appropriately a control group. In the case of CE young people should be either at school (and in that case they cannot be compared because at the end of CE courses they are still at school) or in training. In fact traditional counterfactual evaluation doesn’t fit those situations where the policy covers compulsorily all the eligible population.

A final element which left us unsatisfied is linked to the variable describing the employment outcome, because here again the placement is a multidimensional phenomenon.

The Structural Equation Modelling (SEM) appeared as an approach able to provide reliable descriptions of latent multidimensional variables, expressed as systems of Composite Indicators (CI), at the input/output/outcome levels. Previous analyses (probit models) failed in describing the initial disadvantage of young and adult trainees, who generally come from difficult familiar situations, hard social background, low educational attainments. In particular, family variables, endowment variables and individual network ones did not prove significant in probit models of employability (Nosvelli 2012, Benati et al., 2014 a and b). Moreover SEM adopts a systemic view, able to assess the relations among latent multidimensional variables. This is very important in the case of CE, where the solution adopted by education evaluators to assess the impact is the use of Value Added approach (Wainer, 2004; Lissitz, 2005; OECD, 2008), which implies a comparison between entry and exit conditions.

2. SEM METHODOLOGY AND THE MIXED TWO-STEP APPROACH

The construction of a CI implies the search for a suitable synthesis of a number of observed or Manifest Variables (MVs) in order to achieve a simple representation of a multidimensional phenomenon. Accordingly, a CI can be considered as a latent concept, not directly measurable, whose estimation can be obtained through the values of MVs.

The existing literature offers different alternative methods in order to obtain a CI. Structural Equation Modeling (SEM), and specifically the Partial Least Squares approach to SEM (PLS Path Modeling, PLS-PM) can be used to compute a system of CIs. According to this methodology, it is possible to define a CI as a multidimensional LV not measurable directly and related to its single indicators or MVs by either a reflexive or formative relationship or by both (this defines the measurement or outer model). Each CI is related to other CIs, in a systemic vision, by linear regression equations specifying the so called Structural Model (or Inner Model).
The choice of using SEM as the methodological framework is useful for several reasons, particularly: (i) the possibility of obtaining, simultaneously and coherently with the estimation method, a ranking of individuals for specific indicator; (ii) the possibility of comparing systemic indicators in space and in time; and (iii) the possibility of estimating the hypothesized relationships without making assumptions about data distribution.

Two different approaches exist to estimate model parameters in SEMs: the Covariance-Based (Jöreskog, 1978) techniques and the Component-Based techniques (Wold, 1982). PLS-PM approach to SEM has been proposed as a Component-Based, where the LVs (i.e CIs) estimation plays a main role. As a matter of fact, the aim of Component-Based methods is to provide an estimate of the LVs in such a way that they are the most strongly correlated with one another (according to the path diagram structure) and the most representative of each corresponding block of MVs.

PLS-PM is a suitable tool for the investigation model with a high level of abstraction, in cases where the building of a system of CIs depends on different levels of construction. Higher-Order Constructs are explicit representations of multidimensional constructs that exist at a higher level of abstraction and are related to other constructs at a similar level of abstraction completely mediating the influence from or to their underlying dimensions (Chin, 1998). In Wold's (1982) original design of the PLS-PM, it was expected that each construct would be necessarily connected to a set of observed variables. On this basis, Lohmöller (1989) proposed a procedure to treat hierarchical constructs, the so-called hierarchical component model.

The hierarchical constructs or sayings are multidimensional constructs that involve more than one dimension and we can distinguish them from the one-dimensional constructs that are characterized by a single underlying dimension. There are two main approaches existing in the literature: the Repeated Indicators Approach and the Two Step Approach. The Repeated Indicators Approach (Lohmöller, 1989) is the most popular approach when estimating Higher-Order Constructs in a PLS-PM (Wilson, 2009). The procedure consists of taking the indicators of the Lower-Order Constructs and using them as the MVs of the Higher-Order LV.

The Two-Step Approach is divided in two phases. In the first step the LV scores of the lower-order constructs are computed without the Second-Order Construct (Rajala et al., 2010). Then, in the second step, the PLS-PM analysis is performed using the computed scores as indicators of the Higher-Order Constructs.

In the case where a Higher-Order Construct is formatively related to the Lower-Order dimensions and each construct is reflectively measured by its MVs, the Two Step Approach works better than the other approach, but each approach presents some limitations. Particularly, one aspect of the Two Step Approach is taken into account: namely, the meaning of component for each Lower-Order Construct. In the classic Two Step Approach, the only first component of the Lower-Order Constructs is estimated without the Higher-Order Construct.
This first component is the one that best represents its block of MVs.

Next, these first components are included in the analysis as indicators of the Higher-Order Construct.

In order to resolve the issue related to the predictive power of the component for each Lower-Order Construct, the Mixed Two Step Approach is proposed by Cataldo (2016).

The Mixed Two Step Approach begins with the implementation of the PLS-PM in the case of the Repeated Indicators Approach. Because the Second-Order Construct has no MVs of its own, it is considered as formed of all the MVs of the First-Order Constructs.

Starting from this structure, a PLS-PM algorithm is performed in such a way as to obtain the scores of each block.

Once the scores for the blocks have been obtained, these will be the MVs of the Second-Order Construct. At this moment, once that scores of the PLS-PM are assigned as indicators of the Second-Order Construct, the PLS-PM algorithm can be implemented (Fig. 1).

Therefore, this method is proposed in order to use the component that is the best representative of its block and, at the same time, has the best predictive power on the Higher-Order LV.

3. A SEM APPROACH TO EVALUATE INITIAL TRAINING

This study applies the SEM methodology to the case of initial training in compulsory education in Piedmont, a region in North-West Italy. In 2013, about 4,053 young
students successfully completed a pluriannual VET course in that region.

The SEM approach aims at evaluating VET impact in enhancing trainees' human capital and, thus, employability.

3.1 The model

Aiming at conceiving a policy evaluation strategy capturing eventual enhancement in individual employability due to VET, the structural model in fig. 2 has been developed. It consists of 17 LVs, describing the multidimensionality of human capital (HC from now on) and employability in the compulsory education context. In particular, HC is disentangled in both an input (HCI) and an output (HCO) components, in order to evaluate the direct impact of VET.

In similar studies applying the SEM approach, Dagum and colleagues (Dagum and Slottje, 2000; Vittadini et al., 2003; Dagum et al., 2007) define an household HC as that multidimensional non-observable construct generated by personal ability, investments in the education, and home and social environments.

Partially adapting this model to the VET case study, individual HCI is a first-order hierarchical LV constructed by three components: social background, individual endowment, and previous education.

Then, HCI is enhanced by VET, and the resulting HCO directly influences employability, which is the hierarchical outcome, along with individual job search strategies and the use of job services. This outcome block resembles De Battisti et al. (2014) model, where personal employability is defined as the ability to identify and realize career opportunities and it is formed by social capital, proactivity, and self-efficacy. In our case, HCO echoes social capital, which has both a professional and a familiar components, while proactivity is described by both job search strategies and the use of job services. In fact, there is evidence that both components enhance trainees’ employment probability (Sella, 2014).

Concerning measurement models, table 1 illustrates all MVs and their afferent first-order construct. Social background is defined by information about both the birth and present families -which are generally the same for compulsory education students-, and about the living neighborhood.

First of all, the birth family block is justified by the evidence that educational disadvantage can perpetrate itself over generations, and family background can change the effects of an equal level of education (Wössmann, 2003).

Hence, the corresponding measurement model describes parents’ education and job. Secondly, the present family block reflects the effect that family characteristics have on HC investment. Following Dagum et al. (2007) and considering the lack of truly quantitative information on household wealth, the block is measured by students’ perception about their familiar, economic, and housing conditions.

Finally, students’ judgment about their living neighborhood (safety, cleanliness, services, greenness, means of transport) is inspected as a proxy for socio-economic disadvantage, due to its role in youths’ HC accumulation (Lauro and Ragazzi, 2011). Regarding the retrospective approach (Jorgenson and Fraumeni, 1989; Le et al., 2005), HC value is measured as well by its production costs, demand, and nonmarket
activities. Hence, the individual background is explored, in the form of personal endowment (internet, phone, means of transport) and spending power (consumption of events, sport, restaurants). Finally, educational indicators are considered (Dagum et al., 2007), that complete the HCI block and describe students’ background at their VET enrolment. The VET policy block, which augments HCI, is measured by the different cores of the policy, i.e. its ability to provide support to students’ needs, its promotion of school-to-work transitions, its teaching and practical contents (Benati et al., 2013).

Finally, the outer model for job search strategies characterizes students’ effort in different job search channels (communication media, job referrals, employment services, employees), while the model measuring students’ use of job services describes how that particular service (vocational guidance, expert advice, mentoring) was effective during their job search.

3.2 Survey data

Data are collected by Computer-Assisted Telephonic Interview methodology, surveying a representative sample of 622 young trainees who successfully attended VET during 2013. They represent the 15.3% total initial VET trainees, stratified by nationality and gender (see Ragazzi et al., 2015 for full details). Students’ micro-data come from regional monitoring and administrative archives. The questionnaire has been developed such that all MVs are ordinal Likert scales (see tab. 1).

<table>
<thead>
<tr>
<th>First-Order Construct</th>
<th>MVs</th>
<th>Loadings</th>
<th>First-Order Construct</th>
<th>MVs</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth family</td>
<td>Parents' education</td>
<td>0.502</td>
<td>Personal support</td>
<td>Reception</td>
<td>0.812</td>
</tr>
<tr>
<td></td>
<td>Father's job</td>
<td>0.685</td>
<td>Individual problems</td>
<td>0.690</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mother's job</td>
<td>0.790</td>
<td>Learning problems</td>
<td>0.746</td>
<td></td>
</tr>
<tr>
<td>Family life conditions</td>
<td>Familiar situation</td>
<td>0.764</td>
<td>Personal relations</td>
<td>0.674</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economic conditions</td>
<td>0.835</td>
<td>Teaching</td>
<td>0.650</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Home</td>
<td>0.810</td>
<td>Laboratories</td>
<td>0.710</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Safety</td>
<td>0.784</td>
<td>Pedagogy</td>
<td>Traineeship</td>
<td>0.703</td>
</tr>
<tr>
<td></td>
<td>Cleanliness</td>
<td>0.768</td>
<td>Business tutor</td>
<td>0.751</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Means of transport</td>
<td>0.692</td>
<td>School tutor</td>
<td>0.807</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>0.715</td>
<td>Labour market insertion</td>
<td>0.848</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenness</td>
<td>0.641</td>
<td>Usefulness for work</td>
<td>School-to-work transition</td>
<td>0.843</td>
</tr>
<tr>
<td>Endowment</td>
<td>Means of transport</td>
<td>0.445</td>
<td>Labour services</td>
<td>0.768</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Telephone type</td>
<td>0.730</td>
<td>Job services</td>
<td>Vocational guidance</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td>Internet connection</td>
<td>0.746</td>
<td>Expert advice</td>
<td>0.760</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Restaurant</td>
<td>0.747</td>
<td>Mentoring</td>
<td>0.743</td>
<td></td>
</tr>
<tr>
<td>Spending power</td>
<td>Shows</td>
<td>0.616</td>
<td>Job search strategies</td>
<td>Means of communication</td>
<td>0.817</td>
</tr>
<tr>
<td></td>
<td>Sports</td>
<td>0.615</td>
<td>Job referrals</td>
<td>0.804</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holidays</td>
<td>0.514</td>
<td>Traineeship</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>Grade at enrolment</td>
<td>0.640</td>
<td>Employment services</td>
<td>0.787</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Performance at school</td>
<td>0.875</td>
<td>Previous employers</td>
<td>0.755</td>
<td></td>
</tr>
</tbody>
</table>
3.3 Estimation results

The above theoretical model for initial training has been estimated by the Mixed Two-Step approach to SEM, using the “plspm” package in R statistical software (Sanchez and Trinchera, 2012). All measurement blocks are reflexive (mode A), due to algorithm requirements in repeated estimation.

Tab. 1 presents the loadings for each MV, i.e. the OLS coefficient of a simple regression of the MV on its LV, that describes how the MV reflects the corresponding LV. Blocks unidimensionality, which is fundamental in reflexive models, has been verified by Principal Component Analysis, guaranteeing that all included MVs reflect a single latent concept. Commonality indices in fig. 2, that assess measurement models reliability, reveal that both personal endowment and spending power are quite poorly represented. On the contrary, the high R2 indices show a good predictive power for all structural relations, while redundancy indices on endogenous blocks show a slightly lower goodness for the inner model describing HCI, and an higher goodness for the VET inner model. In any case, the overall model performs quite well (Goodness of Fit index = 0.72).

![Fig. 2 Path coefficients and goodness indices of a SEM model to evaluate the role of initial training policies on individual employability.](image-url)
Analysing path coefficients in fig. 2, it emerges a minor role of parents’ education and job (birth family block, 0.30) in assessing students’ social background, while both the perceived quality of family life conditions (0.53) and local environment (0.52) play a stronger role. In turn, social background is the construct with the highest impact (0.66) on HCI at enrolment, while the low impact of previous education (0.40) is justified by both the fact that the block contains just few MVs and that all initial training students have a standard middle school diploma. On the contrary, the VET block is much important (0.68) in determining HCO, which is in turn the major determinant (0.58) of hierarchical individual employability, i.e. our output construct. On the contrary, the role of other active labour policies (labour services, 0.34) is quite restrained, while job-search strategies show an interesting impact (0.47).

4. CONCLUSIONS

Policy evaluation is a completely new field of application (as far as we know) for SEM composite indicators. This implies that we had not the possibility to follow the traces of previous researchers in designing our model. With this exercise there has been a long and valuable learning process in the definition of correct manifest variables. Starting from this experience, for non statisticians (evaluators, policymakers) wanting to undertake the SEM journey it must be pointed out that this methodology demands properly designed survey data. Our preliminary trials have shown clearly that it is very hard to adapt data collected for different purposes.

This initial work has nevertheless proved the enormous potentialities of SEM approach also for policy evaluation. In fact, with this model we succeeded in:

- Addressing very well input multidimensionality. We were able to characterise and measure the multidimensional latent variables impacting human capital when enrolling at vocational training (education, social background, individual background).
- From a practical perspective this implies the possibility to develop and employ this model for individual profiling (in view of career guidance and of policy customisation).
- From a policy perspective this implies the possibility to employ the model to better design training policies, complementing them with actions aiming at the recovery of the most striking aspects of disadvantage (critical area variables: high impact and low mean).
- We had the possibility to appreciate the added value of labour policies and, in particular, of training services, in a context where it is impossible to find a proper control group and consequently, to assess the net impact via counterfactual methods. Anyway it must be pointed out that in this exercise SEM reliability in evaluating the net impact of vocational training is limited by the low quality of indicators describing VT characteristics.

At the present stage of our research work we were not able to address outcome multidimensionality. The development of the outcome block, describing individual
integration in the labour market, will be the object of future research work. By the way this will allow further reflection upon pre- and post-treatment human capital. Anyway we think that this extension will be better performed on a different sample, concerning adult trainees. This because the very young age of trainees included in the sample adopted for this model causes huge placement difficulties, regardless any difference in background.

In conclusion SEM shows many potential advantages in

- measuring both pre- and post-treatment characteristics
- in evaluating differences between treated and not treated populations

Taking into account that it is impossible, for practical computational reasons, to use the same model for treated and nontreated individuals, we think that - where traditional counterfactual methods are applicable - SEM should not be seen as an alternative rather as a complementary approach.
5. REFERENCES


Benati I, Ragazzi E (ed), Santanera E, Sella L (2014b) Gli esiti occupazionali delle politiche formative in Piemonte. 2° Rapporto annuale di placement. CNR-Ceris e Regione Piemonte, Torino


the literature (No. 05/10). New Zealand Treasury.


