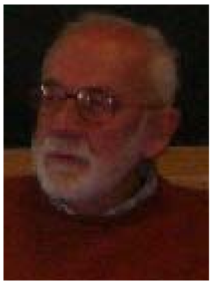
A TWO-LEVEL STRUCTURAL EQUATION MODEL FOR EVALUATING THE EXTERNAL EFFECTIVENESS OF PHD

Bruno CHIANDOTTO[1]
PhD, University Professor,
Department of Statistics "*G. Parenti*"
University of Florence, Italy

E-mail: chiandot@ds.unifi.it



Lucio MASSERINI[2]
PhD, Applied Statistician,
Statistical Observatory,
University of Pisa, Italy

E-mail: lucmas@ec.unipi.it

Abstract: *In recent years the number of PhDs in Italy has significantly grown and purposes of PhD courses have expanded from the traditional ones. The analysis of the contribution of PhD title for job placement and employment condition of PhDs is an important tool for evaluating the quality and the effectiveness of PhD courses. For this reason, knowledge of the employment status and career of PhDs becomes essential and can help to reduce the gap between academia and labour market. The aim of this paper is to estimate a two-level structural equation model with latent variables to assess the external effectiveness of PhD. The analysis is performed using data from the research "Current situation and employment prospects of PhDs", commissioned by National Committee for the Evaluation of the University System (CNVSU) to the Department of Statistics "G. Parenti" of the University of Florence. The proposed measure of "external effectiveness" is a latent variable obtained by evaluating the level of satisfaction with the employment status of PhDs who achieved the title in 2008. The opinion was expressed one year after obtaining PhD on a ten ordered point scale. External effectiveness indicators used are Consistency with studies, Utilization of the acquired skills and Compliance with the cultural interests.*

Key words: *structural equation modelling, multilevel analysis, latent variables, external effectiveness*

1. Introduction

PhD was introduced in the Italian university system by the decree of the President of the Republic 382/1980 as third level of higher education and was subsequently ruled by the Law 210/1998 and by the Ministerial Decree 224/1999. The issue of internal evaluation in universities was introduced by the laws 168/89 and 537/93. The first law provides for the implementation of forms of internal control on efficiency and on management results, the second law provides for the establishment of Internal Evaluation Unit (NVI) in the universities.

The evaluation of the quality of the university system can be placed in the scheme proposed by Lockheed & Hanushek (1994) and generalized by Chiandotto (2004, 2008). According to this approach, the estimated overall performance of an educational system can be decomposed into three distinct stages: the first is the assessment of how resources are employed to get the expected result (efficiency analysis); the second is the qualitative assessment of the results and the level of achievement of objectives (effectiveness analysis); the third is the subjective perception of the subjects involved in educational processes. These evaluations can be made with an internal or external perspective, depending on whether we pay attention to results and achievements respectively when the agents are still in the university system, or instead are outside, typically in the labour market.

This paper focuses on the qualitative assessment of the results obtained by PhDs in the labour market and constitutes an analysis of external effectiveness. More specifically, external effectiveness is analysed by evaluating the level of satisfaction with the current employment status with respect to a set of attributes.

2. Analysis of the external effectiveness

The external effectiveness of a training process can be considered as the contribution of its typical elements, in terms of knowledge and acquired skills, to the individual success in the labour market, net of individual, economic and environmental factors.

In general, there is no well recognized measure of external effectiveness because it is an abstract construct that cannot be directly measured. For this reason we propose to use a set of observed indicators that capture different aspects of the external effectiveness and can reasonably be considered related to this. So, the external effectiveness of PhD is treated as a latent variable, whose values are not observable (Jöreskog & Sörbom, 1979) but whose existence determines association between the observed values of the considered indicators, that otherwise would be uncorrelated (conditional or "local" independence property).

The proposed measure of external effectiveness is based on the evaluation of the satisfaction level of PhDs over a set of features about the employment status one year after achieving the PhD title. The level of satisfaction is evaluated on a ten ordered points scale where 1 = "not at all" and 10 = "very much". Each score is considered as an indicator of an underlying latent variable, whose values are expressed on a continuous scale observable only with a categorical response variable through a set of threshold parameters.

The proposed external effectiveness indicators are: *Consistency with studies*, *Utilization of the acquired skills* and *Compliance with the cultural interests*. In practice, other factors being constant, the PhD title is considered effective respect to the current employment status if the activity is consistent with studies carried out, if the acquired skills are actually used and whether the work is in line with the cultural interests of PhDs.

The effectiveness of PhD respect to the current employment status, measured in terms of satisfaction with the considered indicators, represents the endogenous variable analyzed and is defined *External effectiveness*, denoted by (*E*).

The hypothesized relationship between the latent variable *External effectiveness* and the observed indicators (measurement model) is defined by a confirmatory factor model (Jöreskog, 1969) and is graphically represented as follows (Figure 1).

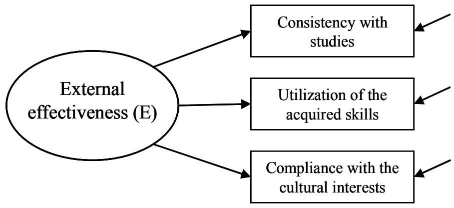


Figure 1. Relationship between the latent variable "*External effectiveness*" and the observed indicators.

The ellipse represents the latent variable while the rectangles represent the observed indicators. The straight directional arrows starting from the latent variable represent the relationship between the latent variable and the observed indicators while the arrows pointing to the observed indicators represent the measurement errors.

The analysis is performed using a two-level structural equation model with latent variables, where the PhDs are considered as first level units and the PhD courses as second level units.

The choice of a two-level analysis is justified to explain the hierarchical data structure and to analyse in a single statistical model variables measured at different levels. In this case, the PhDs are clustered in PhD courses and if they share unobserved factors we may expect related values of the response variables. As a result, observations may not be independent and identically distributed. In this case, a two level analysis allows to get accurate parameters and standard errors estimates and to explain the variability associated with each hierarchical level. The analysis is limited to only two levels of aggregation mostly because we consider that the main effects on external effectiveness of PhDs are exercised by the units to the next higher hierarchical level (PhD courses).

The choice of a structural equation model is justified by the ability of analyzing multivariate relationships between variables (exogenous and endogenous) through a system of linear equations. This allows to model the same variable as dependent variable in an equation and as independent variable in another equation and to decompose the effects of a variable on other variables in direct and indirect effects. The choice of a structural equation model is also justified by the ability of modelling latent variables, such variables can be only measured indirectly by a set of observed indicators.

The external effectiveness of PhD is influenced by a number of variables, some of which are directly measurable, with or without error (manifest variables), others are not directly measurable (latent variables). Among the manifest variables we consider the demographic characteristics (age and gender), variables able to depict the PhD experience (scholarship, a period of study spent abroad and job activity during PhD), a variable representing the actual job (satisfaction level with the current salary), a variable able to define the quality of PhD courses, as the % of those who would enrol again to the same course (*Re-enrollment* aggregated for each PhD course), and finally, the subject area of PhD course. However, for our research purposes, the most challenging aspects are constituted by the latent variables, the only able to represent the effects of complex and multidimensional aspects on the *External effectiveness*. These include the level of satisfaction with the current job, the prospects and the expectations for career, the level of involvement in the job, the job environment and perhaps others. In this research, we consider the latent variables *Satisfaction* with the current job, *Expectations* (prospects and career) and *Participation*, as a measure of involvement in job activities.

These variables are defined in a similar way as done for the *External effectiveness*, evaluating the satisfaction level of PhDs over a set of indicators on a ten ordered points scale where 1 = "not at all" and 10 = "very much". Such indicators are assumed to be correlated because of to the presence of the latent variables. The relationships between the latent variables and the observed indicators (measurement model) are defined by a confirmatory factor model for each latent variable.

For the latent variable *Satisfaction*, denoted by (*S*), the proposed observed indicators are: satisfaction *Compared to the initial expectations*, satisfaction *Compared to the PhD title* and *Overall* satisfaction. These indicators define not only an overall measure of satisfaction but also a comparative one and is focused on the qualitative rather than on the quantitative aspects.

For the latent variable *Expectation*, denoted by (*A*), the proposed observed indicators are *Prospects for wage increase*, *Career opportunities* and *Stability and job security*. The first two indicators reflect aspects relevant during the entire working life period while the third seems to be important especially at the beginning of career because the entry into the labour market is often characterized by precarious types of contract.

For the latent variable *Participation*, denoted by (*P*), the proposed observed indicators are *Independence and job autonomy*, *Involvement in decision-making* and *Skills acquisition*. The first two indicators are clearly related to job participation, evaluated as the level of involvement in working activities. The third has been considered because we assume that a highly professional job probably imply more interest, greater sharing of activities and then a higher level of involvement.

A total of 5 latent variables, 12 observed response variables and 8 observed explanatory variables are used. The latent variables are:

- ***E*** – External effectiveness
- ***EB*** – External effectiveness of the course
- ***S*** – Satisfaction
- ***A*** – Expectations
- ***P*** – Participation.

The observed response variables are:

y_1 – Consistency with studies

y_2 – Utilization of the acquired skills

y_3 – Compliance with the cultural interests

y_4 – Compared to the initial expectations

y_5 – Compared to the PhD title

y_6 – Overall satisfaction

y_7 – Prospects for wage increase

y_8 – Career opportunities

y_9 – Stability and job security

y_{10} – Skills acquisition

y_{11} – Independence and job autonomy

y_{12} – Involvement in decision-making

Finally, the observed explanatory variables are:

x_1 – Satisfaction with the current salary

x_2 – Males vs Females (male =1, female = 0)

x_3 – Age (up to 32 years = 0, more than 32 years = 1)

x_4 – Period of study abroad (yes = 1, no = 0)

x_5 – Job activity during PhD (yes = 1, no = 0)

x_6 – Private vs public University (private = 1, public = 0)

x_7 – Re-enrollment (% of who would enrol again to the same course)

x_8 – Subject area of PhD courses (Area 1: Mathematics, Physics and Computer Sciences - *reference category* -; Area 2: Chemistry and Earth Science; Area 3: Biological, Medical, Agricultural and Veterinary Sciences; Area 4: Architecture and Civil Engineering, Industrial Engineering and Information; Area 5: Science of Antiquity, Literary Philology and Art History, History, Philosophy, Pedagogy and Psychology; Area 6: Law schools, Area 7: Economics and Statistics, Political and Social Sciences).

The variables from y_1 to y_{12} and the variable x_1 measure the satisfaction, as stated before, the level of satisfaction is evaluated on a ten ordered points scale where 1 = "not at all" and 10 = "very much". Moreover note that the latent variable "External effectiveness" is the only variable measured for the first level units (PhDs) and for the second level units (PhD courses).

3. The two-level structural equation model

The model is estimated following the traditional approach to extending structural equation models to a two-level data structure. This approach formulates separate models for the within clusters and between clusters covariance matrices (Goldstein & McDonald, 1988; McDonald & Goldstein, 1989; Longford & Muthén, 1992; Poon & Lee, 1992; Longford, 1993; Muthén, 1994; Lee & Shi, 2001) and can be used with binary, ordinal categorical, censored or continuous response variables as well as with combinations of such variables. More specifically, we use the limited-information weighted least squares estimation approach proposed by Asparouhov & Muthén (2007) as a direct generalization of Muthén (1984) and of Muthén, du Toit, & Spisic (1997) estimation for single level models[3]. With categorical observed response variables, this approach is preferred with respect to others (Skrondal, Rabe-Hesketh & Pickles, 2004) because it can be used to estimate structural equation models with any number of random effects without increasing the computational time. The main limitations are represented by the possibility of estimating models with only random intercepts without calculating the associated values of the random intercepts for each PhD course and by the inability of considering interactions between variables at different levels.

This approach assumes a two-level data structure with N statistical units clustered in J groups[4] where variability in observed response variables exists at individual and cluster level. In this case, the vector of observed response variables referred to unit i in cluster j is indicated with $\mathbf{y}_{ij}$.

The model is constructed as specified in Muthén (1984), by defining an underlying normally distributed latent variable, y^*_{pij}, for the corresponding *p*-th observed variable, y_{pij}. If the *p*-th observed variable is continuous and normally distributed, the latent variable is observed directly, $y^*_{pij} = y_{pij}$. If the *p*-th observed variable is ordinal categorical, the latent variable is defined by a set of threshold parameters, τ_{pk}:

$$y_{pij} = k \Leftrightarrow \tau_{pk} < y^*_{pij} < \tau_{pk+1},$$

where $k = 0,1,\ldots,K-1$ denotes the observed category while $-\infty = \tau_0 < \tau_1 < \tau_2 < ..... < \tau_{K-1} < \tau_K = +\infty$ are the threshold parameters[5].

The vector of latent variables $\mathbf{y}^*_{ij}$ can be decomposed into two independent and additive vectors of latent variables as described in Muthen (1994):

$$\mathbf{y}^*_{ij} = \mathbf{y}_{bj} + \mathbf{y}_{wij}.$$

More specifically, the vector of latent variables $\mathbf{y}^*_{ij}$ is simply composed of a cluster level effect, $\mathbf{y}_{bj}$, which represents the difference between the second level units (random intercepts), and of an individual effect, $\mathbf{y}_{wij}$, which represents the differences between the first level units within each cluster. The two-level model is obtained by specifying two separate structural equation models, respectively for $\mathbf{y}_{bj}$ and $\mathbf{y}_{wij}$. Here, we define $\boldsymbol{\eta}_{wij}$ and $\boldsymbol{\eta}_{bj}$ as the vectors of latent variables, continuous and normally distributed, whereas $\mathbf{x}_{wij}$ and $\mathbf{x}_{bj}$ represent the vectors of observed explanatory variables. Hence, for the first level units (within clusters model), the measurement model and the latent variables model are respectively specified as follows:

$$\mathbf{y}_{wij} = \boldsymbol{\Lambda}_w \boldsymbol{\eta}_{wij} + \boldsymbol{\varepsilon}_{wij},$$
$$\boldsymbol{\eta}_{wij} = \mathbf{B}_w \boldsymbol{\eta}_{wij} + \boldsymbol{\Gamma}_w \mathbf{x}_{wij} + \boldsymbol{\zeta}_{wij}.$$

For the second level units (between clusters model), the measurement model and the latent variables model are respectively specified as follows:

$$\mathbf{y}_{bj} = \boldsymbol{\nu}_b + \boldsymbol{\Lambda}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\varepsilon}_{bj},$$
$$\boldsymbol{\eta}_{bj} = \boldsymbol{\alpha}_b + \mathbf{B}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\Gamma}_b \mathbf{x}_{bj} + \boldsymbol{\zeta}_{bj}.$$

The parameters to be estimated, respectively for the first level and for the second level units, are the factor loading matrices, $\boldsymbol{\Lambda}_w$ and $\boldsymbol{\Lambda}_b$, the coefficient matrices between the latent variables, $\mathbf{B}_w$ and $\mathbf{B}_b$, the coefficient matrices between the latent and the observed variables, $\boldsymbol{\Gamma}_w$ and $\boldsymbol{\Gamma}_b$, the vector of measurement intercepts, $\boldsymbol{\nu}_b$, the vector of intercept terms for the equations, $\boldsymbol{\alpha}_b$, and the vectors of threshold parameters, $\boldsymbol{\tau}_{pk}$. Furthermore, $\boldsymbol{\varepsilon}_{wij}$ and $\boldsymbol{\varepsilon}_{bj}$ are the vectors of measurement errors while ζ_{wij} and ζ_{bj} are the vectors of disturbances, independent and normally distributed variables with zero mean and full variance-covariance matrices respectively $\boldsymbol{\Theta}_w$, $\boldsymbol{\Theta}_b$, $\boldsymbol{\Psi}_w$ and $\boldsymbol{\Psi}_b$.

The estimation of the structural parameters is performed with a three-stage limited-information procedure as described in Muthén (1984) and in Muthén & Satorra (1995). In the first stage, first order statistics (thresholds, intercepts and regression coefficients) are consistently estimated by maximum-likelihood (ML) as described in Muthén & Asparouhov (2006). In the second stage, second order statistics (correlations or covariances) are consistently estimated by conditional ML for given first stage estimates as described in Muthén & Asparouhov (2006). The parameters obtained in the first two stages of the procedure are then summarized in a vector, denoted by $\mathbf{s}$. The third stage uses a weighted least-squares estimation method to fit the structural model. A key component in this step is the development of a weight matrix corresponding to the asymptotic covariance matrix of the statistics computed in the first two stages, as described in Muthén & Satorra (1995). This matrix is denoted by $\mathbf{W}$ and includes the vector of first derivatives of the elements computed in the first two stages of the procedure. The weighted least squares fit function is the following:

$$F_{WLS} = (\mathbf{s} - \boldsymbol{\sigma})\mathbf{W}(\mathbf{s} - \boldsymbol{\sigma})',$$

where $\mathbf{W}$ has the same dimension as the vector $\mathbf{s}$ and $\boldsymbol{\sigma}$ is the corresponding vector of parameters implied by the model. Minimizing the fit function with respect to the model parameters is the last stage of the estimation process. The weighted least square estimates are the parameter that minimize F_{WLS}. When $\mathbf{W}$ is a diagonal weight matrix, as suggested in Muthèn, du Toit & Spisic (1997), we get a limited-information diagonally weighted-least-squares estimation. A beneficial feature of this approach is that $\mathbf{W}$ need not to be inverted, which can be problematic for large models and/or small samples.

4. The formalization of the model

The two-level structural equation model is composed of two systems of linear equation, the first defines the relationships between the latent variables and the observed indicators (measurement model) whereas the second defines the relationships between the latent variables and between the latent variables and the observed explanatory variables (structural model).

The measurement models for the latent variables *External effectiveness* (*E*), *Satisfaction* (*S*), *Expectations* (*A*) and *Participation* (*P*) reflect the considerations exposed in paragraph 3.

The measurement model for the latent variable *External Effectiveness* considers the indicators *Consistency with studies* (y_1), *Utilization of the acquired skills* (y_2) and *Compliance with the cultural interests* (y_3) and can be expressed as follows:

$$y^{*}_{wy_p ij} = \nu_{by_p} + \lambda_{by_p,EBj}\eta_{bEBj} + \lambda_{wy_p,Eij}\eta_{wEij} + \varepsilon_{by_p,EBj} + \varepsilon_{wp,Eij}\ ;\ p = 1,2,3$$

The above equations show the decomposition of the endogenous latent variable *External effectiveness* into a component due to the PhD course j, represented by the factor loadings $\lambda_{by_p,EBj}$, and into a component due to PhDs i clustered in PhD course j, represented by factor loadings $\lambda_{wy_p,Eij}$.

The measurement model for the latent variable *Satisfaction* considers the indicators *Compared to the initial expectations* (y_4), *Compared to the PhD title* (y_5) and *Overall* (y_6), and can be expressed as follows:

$$y^{*}_{wy_p ij} = \lambda_{wy_p,Sij}\eta_{wSij} + \varepsilon_{wy_p,Sij}\ ;\ p = 4,5,6$$

The measurement model for the latent variable *Expectations* (A) considers the indicators *Prospects for wage increase* (y_7), *Career opportunities* (y_8) and *Stability and job security* (y_9), and can be expressed as follows:

$$y^{*}_{wy_p ij} = \lambda_{wy_p,Aij}\eta_{wEij} + \varepsilon_{wy_p,Aij}\ ;\ p = 7,8,9$$

The measurement model for the latent variable *Participation* (*P*) considers the indicators *Skills acquisition* (y_{10}), *Independence and job autonomy* (y_{11}) and *Involvement in decision-making* (y_{12}), and can be expressed as follows:

$$y^{*}_{wy_p ij} = \lambda_{wy_p,Pij}\eta_{wPij} + \varepsilon_{wy_p,Pij}\ ;\ p = 10,11,12$$

The structural model explicates the two-level data structure by formalizing between and within clusters equations for the latent variable *External effectiveness* (*E*) which is the only two-level variable. As a consequence, we can define a within groups latent variable, *External effectiveness* denoted by (E_w) and a between group latent variable, *External effectiveness of the PhD course* denoted by (E_b). The relationships involving E_w and E_b can be expressed respectively as follows:

$$\eta_{wEij} = \beta_{wE,S}\eta_{S_{ij}} + \beta_{E,A}\eta_{Aij} + \beta_{E,P}\eta_{P} + \gamma_{wE,x_2}x_{x_2 ij} + \gamma_{wE,x_3}x_{x_3 ij} + \gamma_{wE,x_4}x_{x_4 ij} + \gamma_{wE,x_5}x_{x_5 ij} + \gamma_{wE,x_1}x_{x_1 ij} + \zeta_{ij},$$

$$\eta_{EBj} = \gamma_{bEB,x_6}x_{6j} + \gamma_{bEB,x_7}x_{x_7 j} + \gamma_{bEB,x_8}x_{8j} + \zeta_{j}$$

The latent variable *Satisfaction* (*S*) presents a system of relationships expressed in extended form as follows:

$$\eta_{Sij} = \beta_{wS,A}\eta_{A_{ij}} + \beta_{wS,P}\eta_{Pij} + \gamma_{wx_1}x_{1ij} + \gamma_{wx_2}x_{2ij} + \zeta_{ij}$$

The latent variable *Expectations* (A) presents a system of relationships expressed in extended form as follows:

$$\eta_{Aij} = \gamma_{x_1} x_{1ij} + \zeta_{ij} .$$

5. The results

The analysis is performed using data from the research "Current situation and employment prospects of PhDs", commissioned by the National Committee for the Evaluation of the University System (CNVSU, 2010) to the Department of Statistics "G. Parenti" of the University of Florence. The survey population consisted of 9696 units (PhDs) while the respondents were 4223 (43,6%). Because of the low response rate, the results may be affected by the non-response pattern. Nevertheless, the effects of this are not studied here.

This paper analyses the PhDs who achieved the title in 2008 and are currently working (3488 first level units and 1488 second level units). Because of missing data, the final model is estimated considering 3053 first level units (PhDs) clustered in 1323 second level units (PhD courses). As a consequence, some PhD courses are not represented. The average number of PhDs for each course is 2,31 while the median is 2.

The two-level structural equation model is estimated with M*plus* 5.21, using the limited-information weighted least squares estimator indicated with "WLMSV", which returns the robust mean and variance adjusted chi-square goodness of fit. This model presents no convergence problems and is obtained after evaluating the results of different specifications of model arising from the knowledge of the phenomenon of the two authors of this paper.

The results are shown in Tables 1, 2 and 3. These include the non-standardized coefficients estimates (Estimate), the standard errors (Std error), the ratio between the estimates and their standard errors (Estimate/ES), the p-values, the coefficients estimates standardized with respect to the latent variables (Std), the coefficients estimates standardized with respect to the variables y and x (Std YX) and the R-square values. The complete system of relationships is depicted graphically using the notation proposed by Muthén & Muthén (2004) which distinguishes the within groups model (Within), represented in the lower area of the diagram, from the between groups model (Between), represented in the upper area of the diagram (Figure 2). The observed variables are included in rectangles while the latent variables are included in ellipses. In the within cluster model, the solid points in correspondence of the match between the directional arrows departing from the latent variable and the respective observed variables indicate that the relationship between the latent variables and the indicators are cluster level random. In the between-group model, the random intercepts are represented as ellipses because they are latent variables, modelled on a cluster level. The significant relationships are represented with bold arrows (p<0.07). Not significant relationships are represented with dashed arrows.

The model fit indices are acceptable. Although the significant value of the chi-square statistic, $\chi^2 = 294.005; df = 29; p < 0.001$, mainly because of the sensitivity of this statistic to the sample size, the alternative fit indices are encouraging: *CFI = 0.96*, *TLI = 0.95*, *RMSA = 0.005*, *SRMR (within clusters) = 0.11* and *SRMR (between clusters) = 0.74*.

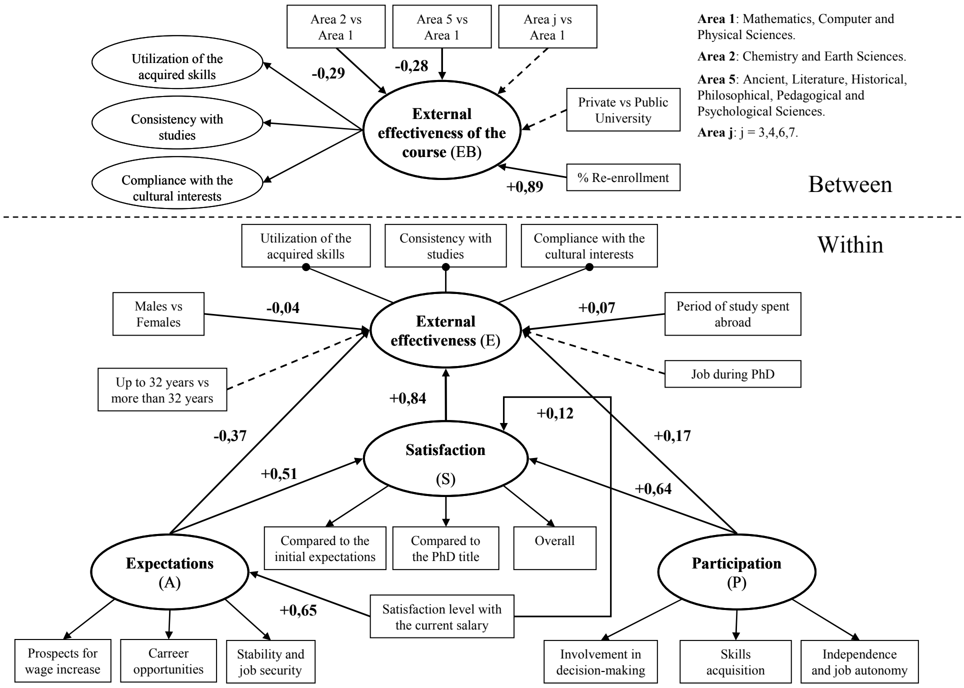


Figure 2. Graphical representation of the proposed model

5.1. Analysis of the two-level data structure

The analysis of the two-level data structure is firstly performed computing the intraclass correlation coefficient for each indicator of the endogenous latent variable *External effectiveness*. The values obtained are the following: *Consistency with studies* $(\rho = 0.02)$, *Utilization of the acquired skills* $(\rho = 0.04)$ and *Compliance with the cultural interests* $(\rho = 0.01)$. The results show a small level of association of the observed responses within each PhD course. As a consequence, the proportion of variability due to the cluster level (PhD course) is very low.

The intraclass correlation coefficient for the endogenous latent variable *External effectiveness* (*E*) is obtained estimating the null model and decomposing the total variance of the latent variable into within and between clusters components. As a result, the intraclass correlation coefficient assumes the following value:

$$ICC_E = \rho_E = \frac{\psi_B}{\psi_B + \psi_W} = \frac{0.619}{19.049} = 0.03$$

Here, $Var(\eta_{Bj}) = \psi_B$ and $Var(\eta_{Wij}) = \psi_W$ are respectively the between and the within clusters variance components of the endogenous latent variable *External effectiveness*. The value obtained, again, is indicative of a moderate level of association in the observed responses within each PhD course. Given the small value of the intraclass correlation coefficient and the small average clusters size, also the design effect assumes a low value with a magnitude less than two. However, in this case it is useful to note that the value of the intraclass correlation coefficient calculated for a latent variable is attenuated compared to the intraclass correlation coefficient calculated for an observed variable. This is due to the presence of the measurement error whose value adds to the residual variance component summarized in ψ_W, increasing the denominator of the ratio.

The hypothesis test for the between-cluster variance component ψ_B is performed using the Likelihood-ratio test where the null and the alterative hypotheses are respectively

$H_0 : \psi_B = 0$ and $H_1 : \psi_B > 0$ [6]. The observed value of the test statistic leads to reject the null hypothesis $(p < 0.001)$. As a consequence, the effect of PhD course on "External effectiveness" is significant, although its value is rather low. This result, together with the possibility of analyzing variables measured at different levels in a single statistical model, justifies the two-level analysis although the level of dependence in the response variables within each PhD course is relatively low.

5.2. Analysis of the measurement models

The values assumed by the factor loadings referred to the measurement models show that the endogenous latent variable *External effectiveness* is well measured by their respective indicators, for each level of the hierarchical data structure (Table 1)[7]. For the within part of the model, the R-squared values associated to the factor loadings assume relatively high values, between 0.740 for *Consistency with studies* and 0.958 for *Utilization of the acquired skills*. For the between part of the model, the values of the factor loading coefficients take values even higher and the associated R-squared are all greater than 0.92.

The measurement models seem adequate also for the other latent variables, although the values of the factor loadings are slightly lower. More specifically, these are rather low for the indicator *Stability and job security* of the latent variable *Expectations*, with an associated R-squared value of 0.648, and for the indicator *Skills acquisition* of the latent variable *Participation*, with an associated R-squared value of 0.421. For both these indicators, the proportion of variability unexplained by the latent variables is fairly high.

Table 1. Measurement model: factor loading estimates

Latent variables and indicators	Estim	Std	Esti	P	StdY	R-
Within Clusters Model (WITHIN)						
External effectiveness (*E*)	-	0.02	28.0	0.00	-	0.696
Consistency with studies (y_1)	1.000	-	-	-	0.86	0.740
Utilization of the acquired skills	2.836	0.54	5.22	0.00	0.97	0.958
Compliance with the cultural	1.062	0.07	14.6	0.00	0.87	0.762
Satisfaction (*S*):	-	0.02	33.3	0.00	-	0.780
Compared to the initial	1.000	-	-	-	0.81	0.668
Compared to the PhD title (y_5)	1.250	0.09	13.6	0.00	0.87	0.759
As a whole (y_6)	1.657	0.16	10.1	0.00	0.92	0.847
Expectations (*A*):	-	0.01	26.7	0.00	-	0.415
Prospects for wage increase (y_7)	1.000	-	-	-	0.95	0.903
Carreer opportunities (y_8)	0.637	0.06	10.0	0.00	0.88	0.790
Stability and job security (y_9)	0.280	0.02	11.1	0.00	0.64	0.421
Job Participation (*P*):	-	-	-	-	-	-
Skills acquisition (y_{10})	1.000	-	-	-	0.91	0.835
Independence and job	0.527	0.05	9.22	0.00	0.76	0.584
Involvement in decision-making	0.444	0.04	10.1	0.00	0.70	0.499
Between Clusters Model (BETWEEN)						
External effectiveness of the course (*EB*):	-	0.30 4	3.28 3	0.00 1	-	0.998
Consistentcy with studies (y_1)	1.000	-	-	-	0.99	0.986
Utilization of the acquired skills	2.805	0.79	3.53	0.00	0.99	0.991
Compliance with the cultural	0.267	0.35	0.75	0.00	0.96	0.926

5.3. Analysis of the first level relationships

The relationships between the latent variables (Table 2) referred to the first level units (PhDs) show that the *Satisfaction* $(\beta = +0.842; p < 0.001)$ and the *Participation* $(\beta = 0.175; p = 0.004)$ have a significant and positive direct effect on the endogenous latent variable, *External effectiveness,* as confirmation of our hypothesis. Instead, the direct effect of the *Expectations* on *External effectiveness* is significant but negative $(\beta = -0.366; p < 0.001)$. The *Expectations* $(\beta = +0.515; p < 0.001)$ and the *Participation* $(\beta = +0.638; p < 0.001)$ also have a significant and positive direct effect on the *Satisfaction*. As a result, the *Expectations* and the *Participation* exert an indirect effect on the *External effectiveness* through the *Satisfaction*. The indirect effect of the *Expectations* on the *External effectiveness* is given by the product of the direct effects of the *Expectations* on the *Satisfaction* and of the *Satisfaction* on the *External effectiveness* $(0.515 \times 0.842 = 0.434)$. Similarly, the indirect effect of the *Participation* on the *External effectiveness* is given by the product of the direct effects of the *Participation* on the *Satisfaction* and of the *Satisfaction* on the *External effectiveness*. The total effect of the *Expectations* on the *External effectiveness* is then given by the sum of the direct and indirect effects and assumes a small value $(-0.366 + 0.434 = 0.068)$. Instead, the total effect of the *Participation* on the *External effectiveness* assumes a high value $(0.175 + 0.537 = 0.712)$.

Table 2. Within clusters model: coefficients estimates

Latent and observed variables	Estimate	Std error	Estimate/ES	P value	Std	StdYX
Regression coefficients between the latent variables						
External effectiveness (*E*) respect to:	-	-	-	-	-	-
Satisfaction (*S*)	1.001	0.107	9.343	0.000	0.842	0.842
Expectations (*A*)	-0.203	0.039	-5.145	0.000	-0.366	-0.366
Participation (*P*)	0.131	0.046	2.861	0.004	0.175	0.175
Satisfaction (*S*) respect to:	-	-	-	-	-	-
Expectations (*A*)	0.240	0.025	9.533	0.000	0.515	0.515
Participation (*P*)	0.402	0.045	8.850	0.000	0.638	0.638
Regression coefficients between the latent variables and the observed explanatory variables						
External effectiveness (*E*) respect to:	-	-	-	-	-	-
Males vs Females (x_2)	-0.121	0.055	-2.216	0.027	-0.072	-0.036
Age (x_3)	0.054	0.067	0.799	0.424	0.032	0.016
Period of study abroad (x_4)	0.241	0.069	3.503	0.000	0.143	0.070
Job activity during PhD (x_5)	-0.044	0.067	-0.640	0.522	-0.026	-0.016
Expectations (*A*) respect to:	-	-	-	-	-	-
Satisfaction with the current salary (x_1)	2.419	0.186	12.995	0.000	0.794	0.648
Satisfaction (*S*) respect to:	-	-	-	-	-	-
Satisfaction with the current salary (x_1)	0.232	0.042	5.504	0.000	0.164	0.124
Males vs Females (x_2)	0.082	0.053	1.538	0.124	0.058	0.019

About the observed explanatory variables, a *Period of study abroad* has a significant and small positive effect on *External effectiveness* $(\beta = +0.070; p < 0.001)$ instead an *Employment during the PhD* produces no significant effect $(\beta = -0.016; p = 0.522)$. Finally, we observe a positive and direct effect of the variable *Satisfaction with the current salary* (x_1) on both the *Expectations* $(\beta = +0.648; p < 0.001)$ and the *Satisfaction* $(\beta = +0.124; p < 0.001)$.

5.4. Analysis of the second level relationships

The analysis of the effects of the explanatory variables on the latent variable *External effectiveness of the course* (Table 3) highlights the significant value of the positive coefficient of the observed variable *Re-enrollment*. This variable, although with caution, can be considered indicative of the *quality* of the PhD course attended. Instead, we observe no significant difference in *External effectiveness of PhD* courses in private universities compared to PhD courses in public universities $(\beta = +0.223; p = 0.168)$.

With regard to the subject area of PhD courses, *Mathematics, Physics and Computer Science* has been chosen as the reference category. Compared to this area, PhD courses in *Chemistry and Earth Science* shows a significant and lower level of *External effectiveness* $(\beta = -0.291; p = 0.022)$. The *External effectiveness* is lower also for the PhD courses in *Science of Antiquity, Literary Philology and Art History, History, Philosophy, Pedagogy and Psychology* even though. For this area the significance level is just over five percent $(\beta = -0.285; p = 0.067)$.

Table 3. Between clusters model: coefficients estimates

Latent and observed variables	Estimate	Std error	Estimate/ES	P value	Std	StdYX
External effectiveness of PhD course (EB) respect to:						
Private vs Public University (x_6)	0.223	0.162	1.378	0.168	0.555	0.155
Re-enrollment (x_7)	0.009	0.002	5.596	0.000	0.022	0.894
Area 1: Mathematics, Physics and Computer Sciences (Reference category)	-	-	-	-	-	-
Area 2: Chemistry and Earth Science	-0.415	0.182	-2.287	0.022	-1.034	-0.291
Area 3: Biological, Medical, Agricultural and Veterinary Sciences	-0.220	0.149	-1.483	0.138	-0.548	-0.256
Area 4: Architecture and Civil Engineering, Industrial Engineering and Information	-0.130	0.155	-0.843	0.399	-0.325	-0.119
Area 5: Science of Antiquity, Literary Philology and Art History, History, Philosophy, Pedagogy and Psychology	-0.291	0.159	-1.829	0.067	-0.724	-0.285
Area 6: Law Schools	0.207	0.214	0.966	0.334	0.515	0.134
Area 7:Economics and Statistics, Political and Social Sciences	-0.186	0.179	-1.041	0.298	-0.463	-0.145

6. Conclusions

Thirty years after the introduction of PhD, in a period of transition and transformation of the Italian university system, any attempt to evaluate the earlier experience may provide a useful contribution to review the reform process in place. The aim of this paper was to estimate a two-level structural equation model with latent variables to assess the external effectiveness of PhD. The proposed measure of *External effectiveness* is a latent variable obtained by evaluating the level of satisfaction with the employment status of PhDs who achieved the title in 2008. This model explains the *External effectiveness* taking into account the complex data structure where the PhDs are the first level units and the PhD courses the second level units.

The first level relationships highlights the crucial role played by the *Satisfaction* on the *External effectiveness*. PhDs more satisfied for the current employment status compared to the initial expectations, compared to the PhD title and overall are those for which increase the chance of having a job consistent with studies, in which the acquired skills are used and where there is compliance with the cultural interests. On the *Satisfaction* have a direct effect the expectations, the job participation and the level of current salary. The *Expectations* and the *Participation* have also an indirect effect on the *External effectiveness* through the *Satisfaction*. On the *External effectiveness* have a positive direct effect also a period of study abroad and an employment during the PhD.

The second level relationships show that the more effective PhD courses are those in which the PhDs would be willing to enrol again. With regard to the subject area, the PhD courses in *Chemistry and Earth Science* and *Science of Antiquity, Literary Philology and Art History, History, Philosophy, Pedagogy and Psychology* show a significant and lower level of *External effectiveness* than the PhD courses in the reference subject area (*Mathematics, Physics and Computer Science*). Instead, there is no difference in *External effectiveness* between PhD courses in private universities compared to PhD courses in public universities.

The proposed model has a few limitations, some of these are attributable to the analysis method, the others to the available data. About the analysis method limitations, we first observe that the model is restricted to only two levels of aggregation, so it is not possible to distinguish the variability in the response variables attributable to higher levels (i.e. universities). Moreover, the approach is limited to random intercepts but it is not possible to estimate the values associated with each PhD course. Finally, it is not possible to estimate the effects of the second level variables (or context variables) on the first level response variables. With regard to the available data limitations, we must consider the small number of PhDs included in each PhD course, as a result of a general limited number of PhD students (with or without a scholarship).

Finally, it should be noted that the results undoubtedly depend on the measure of external effectiveness used, particularly by the indicators chosen to represent it. Thus, the proposed measure should be considered as a relative external effectiveness measure. Even with this consideration, there is a belief that these results can be of some use in the activity of planning and management of the third level of university education (PhD) in Italy.

References

1. Asparouhov, T. and Muthén, B. **Multilevel modelling of complex survey data**, Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA section on Survey Research Methods, 2006
2. Asparouhov, T. and Muthén. B. **Computationally efficient estimation of multilevel high-dimensional latent variable models,** In Proceedings of the Joint Statistical Meeting in Salt Lake City. ASA Section on Biometrics, 2007
3. Bollen, K.A. **Structural equations with latent variables,** New York: Wiley, 1989
4. Bottai, M. and Orsini, N. **Confidence intervals for the variance component of random-effects linear models,** The Stata Journal 4, Number 4, 2004
5. Chiandotto, B. **Sulla misura della qualità della formazione universitaria,** Studi e Note di Economia, vol. 3, 2004
6. CNVSU (2010). **Condizione attuale e prospettive occupazionali dei dottori di ricerca**, Internal Report. MIUR, Roma, 2010
7. Gutierrez, R.G., Carter, S. and Drukker, D.M. **On the boundary-value likelihood-ratio tests**. Stata Technical Bulletin. sg-160*, 2001*
8. Gutierrez, R.G. and Elo, I. **On the boundary-value likelihood-ratio tests,** The Stata Journal 3, Number 1, 2003
9. Jöreskog, K.J. **A general approach to confirmatory maximum likelihood factor analysis**. Psychometrika, 34, 1969, pp. 183-202
10. Kaplan, D. **Structural equation modelling (2nd ed.)**, Thousand Oakes, CA: Sage, 2009
11. Muthén, B. **A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators**. Psychometrika. 49, 1984, pp. 115-132
12. Muthén, B. **Multilevel covariance structure analysis,** Sociological Methods and Research, 22, 1994, pp. 376-398
13. Muthén, B., du Toit, S.H.C. and Spisic, D. **Robust inference using weighted least squares and quadratic estimating equations in latent variable modelling with categorical and continuous outcomes**. Unpublished manuscript, 1997
14. Muthén, B. and Satorra. A. **Complex sample data in structural equation modeling. In P. Marsden (Ed.)**. Sociological methodology, Washinghton. DC: American Sociological Association, 1995, pp. 216-316
15. Muthén, B. and Satorra A. **Technical aspects of Muthén's LISOCOMP approach to estimation of latent variable relations with a comprehensive measurement model**, Psychometrika. 60, 1996, pp. 489-503
16. Rabe-Hesketh, S., Skrondal A. and Pickles, A. **Generalized multilevel structural equation modelling**. Psychometrika 69 (2), 2004, pp. 167-190

[1] Bruno Chiandotto is full professor of Statistics at the Department of Statistics "*G. Parenti*" of the University of Florence since 1980.He has taught study, research and teaching activities at the University of California at Berkeley, University of Naples, University of Modena, University of Pisa, Scuola Normale Superiore of Pisa and at the European University Institute in Florence.The research topics addressed, both in their theoretical and operational implications, concerned the definition and estimation of linear and nonlinear, simple and multi-level, statistical models, methods and statistical models for multivariate data analysis, identification and estimation of stochastic processes, methods and models for causal analysis, statistical decision theory.The research carried out in recent years has concerned prevalently the Italian university system, in particular, dealt with issues related to the evaluation and monitoring of training processes with the aim of defining a battery of synthetic indicators of performance.

[2] Lucio Masserini is an applied statistician and is responsible of the Statistical Observatory at the University of Pisa. He received his PhD in Applied Statistics at the Department of Statistics at the University of Florence. His interests include structural equation modeling, multilevel analysis, generalized linear mixed models, latent variable modeling, multivariate analysis and survival analysis.His research areas focus on the evaluation of the university system, the transition from the university to the labor market, the satisfaction analysis, the evaluation and quality of services, the marketing research and the biostatistics.He is Corresponding Member of the Italian Society of Statistics since 2006 and Member of the Working Group "Statistics for the evaluation and the quality of services" of the Italian Society of Statistics since 2009.

[3] This method is implemented in Mplus 5.0 or later and is indicated as "WLSM" or "WLSMV". The difference between the two estimators is how the goodness of fit test is computed because both lead to the same coefficient estimates and standard errors but to different robust chi-square measures. "WLSM" uses a mean corrected (first order) chi-square whereas "WLSMV" uses a mean-and-variance corrected (second order) chi-square (Satorra, 1989).

4 It is also assumed that the clusters are a random sample from a population of clusters and that the units are randomly selected within each cluster.

5 A binary response variable can be considered as a special case of an ordinal categorical variable where $K = 2$.

6 The hypothesis discussed here requires special care because the postulated value lies on a boundary of the parameter space. In this case, the likelihood-ratio test does not have the usual central chi-square distribution with one degree of freedom but may be better approximated as a 50:50 mixture of central chi-squares with zero and one degree of freedom (Gutierrez, Carter & Drukker, 2001). Nevertheless, as suggested by Rodriguez & Elo (2003) we use the traditional chi-square distribution with nominal one-degree-of-freedom test, treating the resulting p-value as a conservative approximation. A different solution to this problem was proposed by Bottai & Orsini (2004).

7 With regard to non-standardized estimates (Estimate), the first factor loading for each latent variable is set to one for identification problems.