Working Paper Series, N. 2, June 2015

Comments about the use of PLS path modeling in building a Job Quality Composite Indicator

Giovanna Boccuzzo
Department of Statistical Sciences
University of Padua
Italy

Mario Fordellone
Department of Statistical Sciences
University of Padua
Italy

Abstract: A composite indicator is formed when elementary indicators are compiled into a single index, on the basis of an underlying model of the multidimensional concept that is being measured. The PLS path modeling allows the estimation of composite indicators and the measurement model could be expressed both as formative and reflective.
In this paper we construct a composite indicator of job quality using the PLS path modeling approach and compare results obtained by the formative and the reflective measurement models of the general concept. We observe that the two approaches can give different results. Consequently, we give some suggestions in order to estimate stable and reliable models.

Keywords: composite indicator, PLS path models, formative approach, reflective approach, job quality.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Mario Fordellone
tel: +39 049 827 4124
fordellone@stat.unipd.it
http://homes.stat.unipd.it/mariofordellone/

Comments about the use of PLS path modeling in building a Job Quality Composite Indicator

Giovanna Boccuzzo
Department of Statistical Sciences
University of Padua
Italy

Mario Fordellone
Department of Statistical Sciences
University of Padua
Italy

Abstract: A composite indicator is formed when elementary indicators are compiled into a single index, on the basis of an underlying model of the multidimensional concept that is being measured. The PLS path modeling allows the estimation of composite indicators and the measurement model could be expressed both as formative and reflective.
In this paper we construct a composite indicator of job quality using the PLS path modeling approach and compare results obtained by the formative and the reflective measurement models of the general concept. We observe that the two approaches can give different results. Consequently, we give some suggestions in order to estimate stable and reliable models.

Keywords: composite indicator, PLS path models, formative approach, reflective approach, job quality.

1 Introduction

A composite indicator (CI) is defined as *a mathematical combination of single indicators that represent different dimensions of a concept, the description of which is the objective of analysis* (OECD, 2008). Then, a CI is based on a hierarchical model (Wetzels et al., 2009), i.e. a measurement of the phenomenon of interest computed from the aggregation of several dimensions (latent variables), and each dimension is computed by the aggregation of several elementary indicators (manifest variables). In the literature of CIs almost every month new proposals are published. The proposals often concern specific methodological aspects potentially relevant for the development of CIs and theirs application. Several authors (Henseler et al., 2009; Hair et al., 2011), recently, have proposed structural equation models (SEMs) approach for CIs building. The statistical methodology of SEMs (or path models) studies the real world complexity by taking into account a number of causal relationships among latent variables, each measured by several observed (or manifest)

variables (Bollen, 1998; Bagozzi and Yi, 2012).
The objectives of this paper are (*i*) to analyze the opportunity given by path models of constructing composite indicators and (*ii*) to analyze in which way the relationship between indicators and latent constructs can give different estimation results. In particular, we will distinguish between two different operationalizations of the relationships between latent variables and their observed indicators in path models: reflective and formative approach (Diamantopoulos, 1999; Cenfetelli and Bassellier, 2009; Ringle et al., 2009; Becker et al., 2012). In the reflective approach, the observed variables are considered as being caused by the latent variable; in the formative approach the latent construct is supposed to be formed by its indicators (Sanchez, 2013). However, the CI traditional approach is based on the formative model, i.e. the indicators are considered as the cause of the latent construct. For instance, the Human Development Index (HDI) (Anand and Sen, 1994) is a summary measure of average achievement in key dimensions of human development: (*i*) a long and healthy life (health dimension), (*ii*) being knowledgeable (education dimension) and (*iii*) have a decent standard of living (standard of living dimension). The HDI is the geometric mean of normalized indices for each of the three dimensions.
In section 2 of this paper we present a brief background of CIs and PLS path models; in section 3 we focus the attention on PLS path modeling (PLS-PM) and Non Metric PLS path modeling (NM-PLS-PM); in section 4 we present an example of CI building on toy data through PLS-PM; in section 5 we propose the construction of a CI of *job quality* through different NM-PLS-PM applications and compare our results with those obtained by Boccuzzo and Gianecchini (2015), that have proposed a Job Quality Composite Indicator based on a traditional approach.

2 Composite indicators and path models

The hierarchical structures of CIs and SEMs are similar. In fact, both methods provide a framework for analysing multiple relationships between a set of blocks of variables, supposing that each block of variables is represented by a latent construct. The relationships among the blocks are established taking into account previous knowledge of the phenomenon under analysis (OECD, 2008; Vinzi et al., 2010b; Sanchez, 2013).
In the next subsections we provide a brief background of CIs and SEMs, introducing how a CI can be defined through a SEM.

2.1 Composite indicators background

A CI is usually formed by various dimensions (latent constructs), each one measured through different elementary indicators (observed/manifest variables). For instance, we can measure the *quality of life* with a CI formed by two dimensions (clearly this is a simplification), *social relations* and *health*, measured through their elementary indicators. This typical hierarchical structure of is shown in Figure 1.
In its most simple formulation, a CI is a weighted average of its dimensions, and

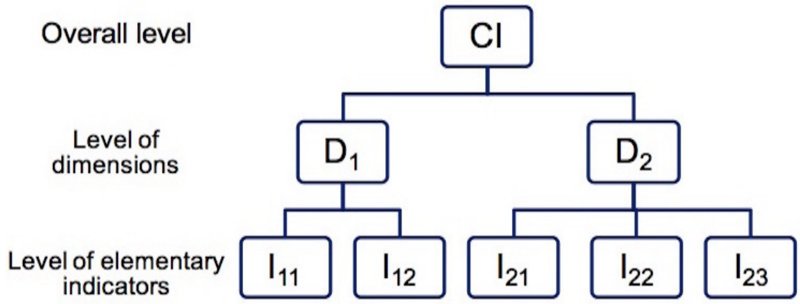

Figure 1: Hierarchical structure of a CI with two dimensions

each dimension is a weighted average of its indicators:

$$CI = \sum_{p=1}^{P} \alpha_p \left(\sum_{j=1}^{N_p} I_j w_j \right), \tag{1}$$

where P is the number of dimensions and p is the index of the dimensions, j is the index of the elementary indicators composing each dimension, N_p is the number of indicators forming the p-th dimension, α_p is the weight of the p-th dimension, I_j is the j-th elementary indicator, w_j is the weight of the j-th elementary indicator. There are different methods for weights construction (OECD, 2008); they are based on statistical approaches, participative approaches or mixed approaches.

2.2 Path models background

SEM is a general term used to describe a family of statistical methods designed to test a conceptual or theoretical model. Some common SEM methods include confirmatory factor analysis, path analysis, and latent growth modeling. The term *structural equation model* most commonly refers to a combination of two things: a *measurement model* that defines latent variables using one or more observed variables, and a *structural regression model* that links latent variables together. The parts of a structural equation model are linked to one another using a system of simultaneous regression equations (Kline, 2011).
Formally, let X be a generic $N \times J$ data matrix, where each row represents a statistical unit described by J observed variables (or manifest variables, MV), and given P latent variables (LV), each described by a group of x_j observed variables, SEM consists (i) in a set of measurement model for the estimation of the P latent variables through the J observed variables and (ii) in a structural model for the estimation of the relationship among the P latent variables (Kline, 2011). In the following, we will define two types of MVs and LVs: first order and second order. The second order LV is the overall latent construct, while the first order LVs are the latent constructs that form the overall latent construct. We define the MVs linked to the first order LVs as first order MVs and the MVs linked to the second order LV as second order MVs. Figure 2 represents the hierarchical structure of a SEM. In the path diagram the boxes represent observed variables, while the circles represent

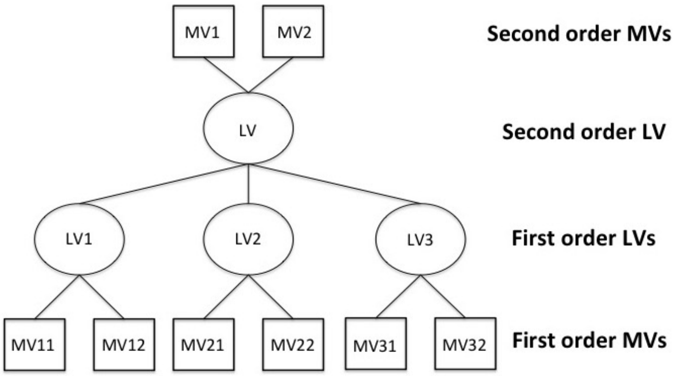

Figure 2: Hierarchical structure of SEMs with three first order LVs

latent variables. Using the CIs terminology, the first order MVs are the elementary indicators, the first order LVs are the dimensions and the second order LV is the composite indicator. However, there is a fundamental difference between CIs and path models structures: in the traditional CI approach the second order MVs are missing, because weights are assigned by a procedure external to the model. On the contrary, in the SEM approach an iterative procedure estimates the relationship between LVs within the model, and to this aim the second order MVs are necessary. Table 1 shows the elements of the SEMs approach and the traditional CIs approach.

Table 1: Elements of SEMs approach and CIs approach

SEMs	CIs
first order LVs	dimensions
second order LV	composite indicator
first order MVs	elementary indicators
second order MVs	not present

In according to Jöreskog and Sörbom (1996), the principal statistical software for the SEMs estimation is LISREL (*LInear Structural RELations*), and his generalizations as SIMPLIS (*SIMPle LInear Structural RELations*), EQS (*structural EQation modeling Software*) and AMOS (*Analysis MOment Structures*). These software use covariance-based techniques (Hsu et al., 2006). However, with the increasing complexity of the theoretical model (e.g. non-linear relations among variables/indicators), researchers have called for new SEM techniques that could address this issue. Recently, component-based technique, with the development of the partial least square (PLS) algorithm (Tenenhaus et al., 2005; Vinzi et al., 2010b; Monecke and Leisch, 2012), have been added to the covariance-based techniques. Furthermore, there is a fundamental problem in the SEMs estimation: in many fields of CIs application as sociology, economy, medicine, etc. there are non-numeric observed variables, e.g. dichotomous, ordinal or likert scale. In this case, the PLS algorithm is not reliable. Russolillo (2012) tried to solve this problem introducing a non-metric partial least square (NM-PLS) method.

3 Formative and reflective approach in path model

Every path model is generally formed by two sub-models: the structural model and the measurement model. The structural model (defined *inner model*) is the part of the model that has to do with the relationships between the latent variables, while the measurement model (defined *outer model*) is the part of the model that has to do with the relationships of a latent variable with its block of manifest variables (Sanchez, 2013). The relationship between a LV and its MVs can be reflective or formative. In the first case, the manifest variables are considered to be the effect of the latent variable, in the second case the manifest variables are considered to be the cause of the latent variable. The choice between formative and reflective model is still an open question.
Wilcox et al. (2008), proposes an extended research review about the use of formative and reflective models. From this review, many different opinions emerge. For instance, Podsakoff et al. (2003) say that *some constructs are fundamentally formative in nature and should not be modeled reflexively*; whereas, Edwards and Bagozzi (2000) assume a behavior overly simplistic suggesting several criteria derived from the literature on causation, that might be employed in this regard. These criteria include association, temporal precedence and the elimination of causal explanations on the data; conversely, Bollen and Ting (2000) note that *establishing the causal priority between a latent variable and its indicators can be difficult* and they offer a promising empirical tool for determining whether the covariance structure among a set of items is more consistent with a formative or reflective measurement model, based on *tetrad analysis*; MacKenzie et al. (2005) suggest that *with respect to the formative model, indicators in reflective measurement model should be highly correlated*. Then, they suppose that the covariance among observed variables will inform the choice of relationship type between measures and constructs.
Many other instances exist. However, according Wilcox et al. (2008), we believe that *the empirical meaning of a formatively/reflectively measured construct depends on the outcome variables in the model and construct's empirical realization will vary from model to model and study to study*. Practically, the conceptualization of the measurement model is often more dependent on the choice of the researcher than some inherent characteristic of a particular construct or observed variables. In fact, while in some cases determining the direction of causation between measures and their construct appears to be easy (Diamantopoulos and Winklhofer, 2001; Jarvis et al., 2003), many instances exist in which a potential indeterminacy from an examination of the items alone may occur and larger research context must be considered. Traditionally, the measurement approach used in path models is the reflective one, i.e. the indicators are considered as the effect of the latent constructs (Diamantopoulos and Winklhofer, 2001), while the CI approach is often based on the formative relationship between elementary indicators and dimensions. Consequently, in this paper we focus the attention on these two alternative measurement ways based on formative and reflective relationships, only for the second order indicators. With respect CI structures, the first order indicators have a formative relationship with its latent constructs. We will analyze the effect that different measurement models have on the final estimates of the global path model. Figure 3 represents an example

of path diagram with formative first order indicators and formative second order indicators, while Figure 4 represents an example of path diagram with formative first order indicators and reflective second order indicators.

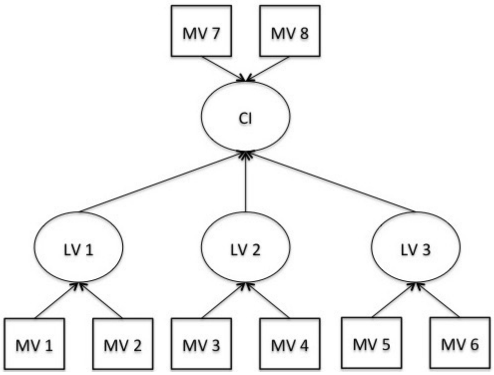

Figure 3: Path diagram with formative first order indicators and formative second order indicators

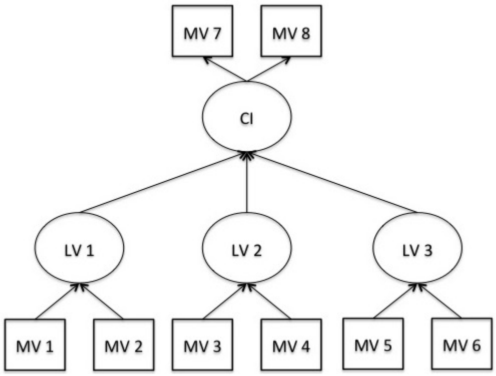

Figure 4: Path diagram with formative first order indicators and reflective second order indicators

We define the path models in Figure 3 and 4 *formative model* and *hybrid model*, respectively (Hauser and Goldberger, 1971; Becker et al., 2012). The formative and reflective measurement are represented by the direction of the arrows. The manifest variables from 1 to 6 are the first order indicators while the manifest variables 7 and 8 are the second order indicators linked to the overall latent construct (composite indicator). Latent variables 1, 2 and 3 are the latent constructs (dimensions) that are linked to the composite indicator. According to Sanchez (2013), the formative measurement model is shown in Equation 2:

$$LV_p = \lambda_p + \sum_{j=1}^{J} \lambda_{jp} MV_{jp} + \epsilon_p, \tag{2}$$

where λ_p is the intercept term of the model, λ_{jp} are the measurement model coefficients (called loadings) and represent the strength and sign of the relations between the response LV_p and the predictors MV_{jp}, while ϵ_p is the residual component of the model. The reflective measurement model is shown in Equation 3:

$$MV_j = \lambda_j + \sum_{p=1}^{P} \lambda_{jp} LV_{jp} + \epsilon_j, \tag{3}$$

where λ_j is the intercept term of the model, λ_{jp} are the measurement model coefficients (called loadings) and represent the strength and sign of the relations between the response MV_j and the predictors LV_{jp}, while ϵ_j is the residual component of the model. In our case $p = 1$. The loadings are calculated as correlations between the latent variable and its indicators. The structural model is shown in Equation 4:

$$LV_{p^*} = \beta_0 + \sum_{p=1}^{P} \beta_p LV_p + \epsilon_p, \tag{4}$$

where β_0 is the intercept term of the model, β_p are the path coefficient and represent the strength and sign of the relations between the response LV_{p^*} (i.e. the composite indicator) and the predictors LV_p, while ϵ_p is the residual component of the model. The hierarchical construct model presented in Figures 3 and 4 can be estimated through PLS path modeling and Non-Metric PLS path modeling. In the next subsections, according to Tenenhaus et al. (2005) and Russolillo (2012), we describe synthetically the steps of the general algorithm in PLS-PM and NM-PLS-PM.

3.1 PLS-PM method

The partial least squares path modeling is a statistical data analysis procedure that is born from the intersection of regression models, structural equation models, and multiple table analysis methods (Vinzi et al., 2010a; Sanchez, 2013). However, in most cases, PLS method is usually referred to as the PLS approach to SEM. Given a $N \times J$ data matrix, where J is the number of MVs that are organized in P blocks, the general algorithm of PLS-PM is structured in three stages (Henseler, 2010; Russolillo, 2012):

1. Get the weights to compute latent variable scores
2. Estimating the path coefficients (inner model)
3. Obtaining the loadings (outer model)

The first stage consists in obtaining the weights that will be used to get the scores of the latent constructs, the second stage is the estimation of the path coefficients of the inner model and the third stage involves the computation of the loadings. We define outer and inner weights with w_{jp} and α_p respectively.
The **first stage** is structured as follows:

- **Step 0**: The iterative process is started by assigning arbitrary values to the outer weights $\tilde{w}_{jp}$.
- **Step 1**: Compute the outer estimation of latent variables as a weighted sum of its respective indicators: $y_{ip} = \sum_{j=1}^{J} x_{ij}\tilde{w}_{jp}$, where y_{ip} is the score of the i-th statistical unit for the p-th latent variable, x_{ij} is the observed value of the i-th statistical unit for the j-th indicator and $\tilde{w}_{jp}$ is the assigned weight of the j-th indicator for the p-th latent variable.
- **Step 2**: Obtain inner weights α_p for each latent variable in order to reflect how strongly the other latent variables are connected to it, i.e. to find the weights for re-calculate the latent variable scores as the linear combination of its associated latent variables. There are three options for determining the inner weights: *centroid*, *factorial* and *path* scheme. The **Centroid scheme** considers the sign direction of the correlations between a LV and its adjacent LVs:

$$\alpha_p = \begin{cases} sign[cor(y_p, y_{p^*})], & \text{if } LV_p, LV_{p^*} \text{ adjacents} \\ 0, & \text{otherwise} \end{cases}$$

where LV_p and LV_{p^*} are two adjacent latent variables, while y_p and y_{p^*} are their scores. The **Factorial scheme** uses the correlation coefficient as the inner weight instead of using only the sign of the correlation. In other words this scheme considers not only the sign direction but also the strength of the paths in the structural model. Then, the inner weights are defined as:

$$\alpha_{jp} = \begin{cases} cor(y_p, y_{p^*}), & \text{if } LV_p, LV_{p^*} \text{ adjacents} \\ 0, & \text{otherwise} \end{cases}$$

where LV_p and LV_{p^*} are two adjacent latent variables, while y_p and y_{p^*} are their scores. Finally, the **path scheme** has the advantage of taking into account both the strength and the direction of the paths in the structural model. However, this scheme presents some problems when the LV correlation matrix is singular.

- **Step 3**: Once the inner weights are obtained, the procedure provides the inner estimation of latent variables. If we define y_{ip} the i-th score of the p-th latent variable LV_p and y_{ip^*} the i-th score of the adjacent latent variable LV_{p^*}, the procedure computes the inner estimation of LV_p as a weighted sum of the adjacent latent variables: $y_{ip} = \sum_{p \neq p^*=1}^{P} y_{ip^*}\alpha_p$.

- **Step 4**: Once the inner estimation is done, the procedure provides to calculate new outer weights as a simple regression of each indicator x_{ij} on its latent score y_{ip}, in the case of reflective indicators, and as a multiple regression of y_{ip} on the x_{ij}, in the case of formative indicators. In matrix notation we have the OLS estimation $(Y_{pj}^{'}Y_{pj})^{-1}Y_{pj}^{'}X_j$ in the case of reflective indicators, while in the case of formative indicators we have the OLS estimation $(X_{jp}^{'}X_{jp})^{-1}X_{jp}^{'}Y_p$, where X is the $N \times J$ data matrix, Y is the $N \times P$ matrix of the latent variable scores, X_j is the $N \times 1$ vector of the j-th indicator and Y_p is the $N \times 1$ vector of the p-th latent variable score.

The final estimates of the latent variables are obtained through the alternation of their outer and inner estimations and the iteration procedure repeats step 1 to step 4 until convergence of the outer weights is achieved.
The **second stage** of the algorithm consists in calculating the path coefficients estimates $\widehat{\beta}_{jp}$ thorough the OLS estimation in the multiple regression on the latent variables. Then, $\widehat{\beta}_p = (Y_p^{'}Y_p)^{-1}Y_p^{'}Y_{p^*}$, where Y_p is the $N \times P$ matrix of variable scores and Y_{p^*} is $N \times 1$ vectors of the adjacent latent variable score.
Finally, the **third stage** of the algorithm consists of calculating the loadings that are usually calculated as correlations between a latent variable and its indicators (Sanchez, 2013): $\lambda_{jp} = cor(X_{jp}, Y_p)$.
Unfortunately, in presence of non numerical variables/indicators, the PLS algorithm is not reliable. Russolillo (2012) tried to solve this problem introducing non-metric partial least square path modeling (NM-PLS-PM).

3.2 NM-PLS-PM method

In PLS techniques the variables are measured on interval or ratio scales. However, in many fields where PLS methods could be applied (e.g. construction of CI), researchers need to analyse models based on variables measured on a non-metric scale, i.e. ordinal or dichotomous. In this subsection we present, according to Russolillo (2012), the methodological proposal to enable PLS techniques to deal with both metric and non-metric variables. Three new PLS-type algorithms have been proposed for the analysis of one, two or several blocks of variables: the Non-Metric NIPALS, the Non-Metric PLS Regression and the Non-Metric PLS Path Modeling, respectively. In this work we focus on the last one (NM-PLS-PM).
The Non-Metric approach in PLS framework is based on the concept of *optimal scaling* (OS).
NM-PLS algorithms represent a new class of PLS algorithms that generalize the standard PLS methods to the treatment of non-metric variables. These methods provide data with a new metric structure, that does not depend on the metric properties of the original data. In other words, NM-PLS methods change non-metric data with a metric, and change metric data with a new metric, making relationships between variables and latent constructs linear, as required in standard PLS models.
Formally, given an observed non-metric variable x^*, e.g. ordinal, in the OS process a numeric (or scaling) value is assigned to each ordered level of x^* such that (*i*) it is coherent with the chosen scaling level and (*ii*) it optimizes the model criterion. To optimize NM-PLS model criteria, for x^*, the correspondent scaling vector must satisfy the following criterion:

$$\underset{\phi}{\text{argmax}}\ cor^2(\tilde{X}\phi, \gamma_{x^*}), \tag{5}$$

where ϕ is the vector of optimal scaling parameters, the matrix $\tilde{X}$ defines a space in which the constraints imposed by the scaling level are respected and γ_{x^*} is an unknown coefficient. Criterion in Formula 5 is optimized by means of the ordinary least squares regression coefficients of γ_{x^*} on $\tilde{X}$, i.e. by projecting γ_{x^*} on the space defined by the columns of $\tilde{X}$. The resulting projection, standardiszd to unitary variance, is the geometric representation of the scaled variable x^*.

4 Example of CI building on toy data

In this section we present an example of CI building through a PLS-PM application on *Spanish Football Data set* (Sanchez, 2013). This data frame contains the results of the teams in the Spanish football league 2008-2009 and consists in 20 observations on 14 variables. The variables (elementary indicators) may be used to construct four latent concepts: *Attack* (ATT), *Defense* (DEF), *Success* (SUCC) and *Indiscipline* (IND). In this case a simple theory is defined: the better the quality of the Attack, as well as the quality of the Defense, give to more Success. While a high-level of Indiscipline leads to less Success. Then the *Success* construct is formed (explained) by *Attack*, *Defense*, and *Indiscipline.*
The 14 elementary indicators are *the total number of goals scored at home*, *the total*

number of goals scored away, *the percentage of matches with scores goals at home*, and *the percentage of matches with scores goals* that are linked to the **Attack** contruct; *the total number of goals conceded at home*, *the total number of goals conceded away*, *the percentage of matches with no conceded goals at home*, and *the percentage of matches with no conceded goals away* that are linked to the **Defense** contruct; *the total number of yellow cards* and *the total number of red cards* that are linked to the **Indiscipline** construct; finally *the total number of won matches at home*, *the total number of won matches away*, *the longest run of won matches* and *the longest run of matches without losing* that are linked to the **Success** construct (Table 2).
For the applications, the R software and *plspm* package have been used.

4.1 Proposal of path modeling

For this application we use the *centroid scheme* for inner weights estimation and we propose two types of path model: the *formative model*, where all the MVs (first and second order) are linked to the corresponding LVs through a formative relationship; and the *hybrid model*, with formative first order indicators and reflective second order indicators. In Figure 5 and Figure 6 the proposed models are shown. Figure 5 shows that the overall latent construct *Success* (CI) is formed by the latent variables *Attack*, *Defense* and *Indiscipline*. The latent variable *Attack* is formed by GSH, GSA, SSH and SSA; the latent variable *Defense* is formed by GCH, GCA, CSH and CSA; the latent variable *Indiscipline* is formed by YC and RC; and, finally, the latent variable *Success* is formed by WMH, WMA, LWR and LRWR. In Figure 6 the latent variable *Success* reflects on WMH, WMA, LWR and LRWR. In the next subsections we present the estimations results and the diagnostic measures of the global model, according to Sanchez (2013).

Table 2: Indicators codification of Spanish Football Data set

Construct	Variable	Description
Attack	GSH	total number of goals scored at home
	GSA	total number of goals scored away
	SSH	percentage of matches with scores goals at home
	SSA	percentage of matches with scores goals away
Defense	GCH	total number of goals conceded at home
	GCA	total number of goals conceded away
	CSH	percentage of matches with no conceded goals at home
	CSA	percentage of matches with no conceded goals away
Indiscipline	YC	total number of yellow cards
	RC	total number of red cards
Success	WMH	total number of won matches at home
	WMA	total number of won matches away
	LWR	longest run of won matches
	LRWL	longest run of matches without losing

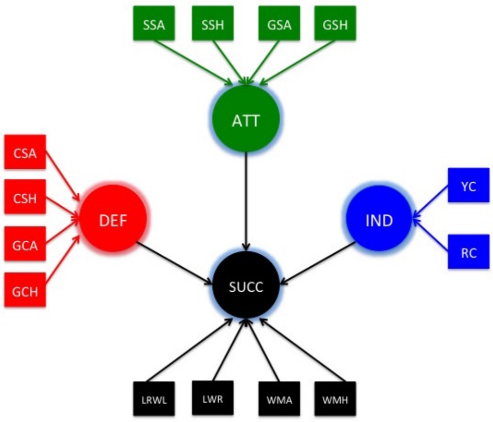

Figure 5: Path diagram with formative model

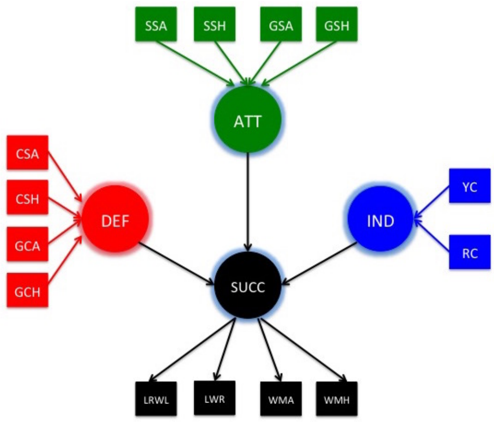

Figure 6: Path diagram with hybrid model

4.2 Results

In the reflective approach, firstly it is necessary to verify the **unidimensionality** of the reflective indicators, that must be in a space of one dimension, represented by the corresponding latent variable. There are three main indices to check unidimensionality: the Cronbach's alpha, the Dillon-Goldstein's rho and the first eigenvalues of covariance matrix (Appendix A for details). The Cronbach's alpha is a coefficient that is intended to evaluate how well a block of indicators measure their corresponding latent construct and a level greater than 0.7 is considered acceptable; Dillon-Goldstein's rho focuses on the variance of the sum of variables in the block of interest. A block of MVs is considered as unidimensional when the Dillon-Goldstein rho is larger than 0.7; finally if a block is unidimensional, then the first eigenvalue of the covariance matrix should be much more larger than 1 and the second eigenvalue should be smaller than 1. Table 3 shows the unidimensionality measures of the hybrid model, applied to the overall latent construct (*Success*).

Table 3: Unidimensionality measures of the hybrid model of *Success*

LV	C.alpha	DG.rho	Eig.1st	Eig.2nd
Success	0.917	0.942	3.22	0.537

In this case Cronbach's alpha is 0.917, Dillon-Goldstein's rho is 0.942 and the first eigenvalue is 3.22, while the second eigenvalue is 0.537. Then we can say that unidimensionality is satisfied. The next step is the analysis of the loadings and the communalities (Appendix B for details) that are contained in the **measurement model**. The loadings are the correlations between a latent variable and its indicators, while the communalities are the squared loadings and are used to measure the part of the covariance between a latent variable and its indicator. Loadings greater than 0.7 and communalities greater 0.5 are considered acceptable. In Table 4 we notice that the hybrid model presents only loadings of SSA, GCH and RC smaller 0.7; while the formative model presents only loadings of SSH, GCH, CSA and RC smaller 0.7.

Table 4: Estimates of hybrid and formative models

	Hybrid model		Formative model	
MV	Loading	Communality	Loading	Communality
		Attack		
GSH	0.955	0.911	0.890	0.791
GSA	0.807	0.652	0.830	0.690
SSH	0.741	0.549	0.617	0.381
SSA	0.658	0.432	0.704	0.496
		Defense		
GCH	-0.424	0.180	-0.546	0.299
GCA	-0.849	0.720	-0.867	0.752
CSH	0.812	0.660	0.762	0.581
CSA	0.802	0.643	0.658	0.433
		Indiscipline		
YC	0.896	0.804	0.890	0.793
RC	0.539	0.290	0.550	0.303
		Success		
WMH	0.916	0.839	0.793	0.629
WMA	0.922	0.850	0.951	0.905
LWR	0.930	0.864	0.815	0.664
LRWL	0.926	0.858	0.980	0.961

After assessing the quality of the measurement model, the next step is to assess the coefficients of the **structural model**. Table 5 shows the estimation of the path coefficients in the structural model, for both formative and hybrid approach.

Table 5: Path coefficients estimates of the structural model, hybrid vs formative model

	Success - Hybrid model			Success - Formative model		
LV	Estimate	Std. Error	t-value	Estimate	Std. Error	t-value
Attack	0.761	0.088	8.690	0.740	0.086	8.650
Defense	0.226	0.089	2.550	0.248	0.087	2.850
Indiscipline	-0.140	0.078	-1.800	-0.146	0.077	-1.900

In both models, only the coefficients of *Indiscipline* are not significantly different from zero. The structural model estimations show that the latent variables *Attack* and *Defense* have a positive and significant relationship with the LV *Success*.
The last step is the examination of the **overall fit measures** R^2 and GoF (Appendix B for details). R^2 indicates the amount of variance of the endogenous latent variable explained by its independent latent variables, while GoF assess the overall prediction performance of the model. It is calculated as the geometric mean of the average communality and the average R^2 value. In the hybrid model, $R^2 = 0.914$ and $GoF = 0.777$, while in the formative model $R^2 = 0.919$ and $GoF = 0.754$.
Observing the inner and outer models results we can say that the two approaches (formative and hybrid) show similar results. Furthermore, in both cases the local and global measures are acceptable. In the inner models, the estimated path coefficient are almost equal.

5 Construction of a composite indicator of job quality

In this section we propose a *Job Quality Composite Indicator* (JQCI) computed by different applications of NM-PLS-PM (Subsection 3.2). Data belong to the Agorá longitudinal survey on the career outcomes of graduates from the University of Padua (Fabbris, 2012). Respondents were interviewed after 6, 12 and 36 months from graduation, using a CATI (Computer Assisted Telephone Interview) technique. Workers were required to give a wide range of information about their current job, the job search activities, the perception of skill and educational mismatch, and the evaluation of their educational program. The survey does not consider people continuing their studies.

According to Boccuzzo and Gianecchini (2015), the relevant dimensions of job quality for young graduates are: *Economic* (ECO), concerning all the aspects that are related to the economic exchange between the worker and the employer, and that are generally included in the formal employment contract; *Professional* (PRO), in which the job characteristics are related to the worker's skills and ambitions; *Work-life balance* (WLB), involving those aspects that affect both worker's personal life and work relationships.

5.1 Data

The data set is composed by 2230 observations and 21 variables (elementary indicators): 10 represent the job conditions of the worker and 11 represent the job satisfaction. In our case, the job conditions are the first order indicators, while the job satisfactions are the second order indicators. Tables 6 and 7 show the details of the first and second order indicators.

5.2 Proposal of path modeling

In the theoretical framework of the JQCI (Boccuzzo and Gianecchini, 2015) the ECO dimension is formed by *hourly wage* and *contract stability*, the PRO dimension is formed by *degree of specialization*, *coherence of the work with the studies*, *supervision of team-work*, *attended career perspectives*, *working in team* and *exploited professional skills*, finally the WLB dimension is formed by *distance between home and work* and *number of working hours in a week*. We remember that, with respect to the traditional CI construction, the path model contains also the measurement model of the general concept (i.e. the job quality), then, the choice of the relationship type used to link the overall latent construct (composite indicator) with its observed variables (second order indicators) is fundamental.

In Section 3 we have discussed about the choice between reflective and formative approach yet. With regard the concept of job quality, the choice between the two approaches is not trivial: *do the job satisfaction facets compose the job quality?* or, conversely, *does the job quality influence the job satisfaction facets?*.

The traditional SEM approach typically utilizes the reflective measurement, even if the use of the formative approach has been recently adopted, especially after the development of PLS models (Bollen, 1998; Law and Wong, 1999; Hsu et al., 2006; Bagozzi and Yi, 2012).

Table 6: Elementary indicators of the job quality data set

Variable	Description	Type	Construct
	Job conditions:		
A1	Hourly wage	numeric	ECO
A2	Contractual stability	ordinal	ECO
B1	Degree of specialization	ordinal	PRO
B2	Coherence of the work with the studies	likert	PRO
B3	Supervision of team-work	dichotomous	PRO
B4	Attended career prospective	dichotomous	PRO
B5	Working in team	dichotomous	PRO
B6	Exploited professional skills	ordinal	PRO
C1	Distance between home and work	ordinal	WLB
C2	Number of working hours in a week	numeric	WLB
	Satisfaction far:		
js1	Stability of the job	likert	JQ
js2	Professionalism acquisitions	likert	JQ
js3	Social prestige derived by the job	likert	JQ
js4	Correspondence between cultural interest and job	likert	JQ
js5	Social utility of the job	likert	JQ
js6	Personal autonomy in the job activity	likert	JQ
js7	Flexibility times of the job	likert	JQ
js8	Free time out of the job	likert	JQ
js9	Place of employment	likert	JQ
js10	Earnings perspectives	likert	JQ
js11	Career perspectives	likert	JQ

Most of the job satisfaction literature embraces the formative approach (Edwards, 2010; MacKenzie et al., 2005): job quality is formed by its different facets, like pay, career opportunities, autonomy.

However, according to James and Jones (1980), the formative approach is questionable. In fact, the overall construct of job satisfaction is conceptually related to other different aspects (expectations, personal capabilities, job context, etc.) that could be linked to the construct in both direction, formative and reflective. Then the use of formative or reflective measurement models is not easy to be chosen and, in our case, it might be appropriate to apply both approaches.

In this case, we try to understand if several measurement models bring to similar or different results. In particular, we have compared the formative (Figure 7) and hybrid (Figure 8) approaches applied to the relationship between the Job Quality latent construct and its second order indicators.

5.3 Results

Table 8 shows the unidimensionality measures of the hybrid model, applied to the overall latent construct (*Job Quality*): we can say that unidimensionality is satisfied. In particular, Cronbach's alpha presents a value over 0.7 (0.783), as well as Dillon-Goldstein's rho with a value equal to 0.836; finally, the first eigenvalue of the covariance matrix is greater than 1 (3.76).

Table 7: Description of the elementary indicators

Variable	Level
A1	Monthly net salary / Monthly working hours
A2	1. Permanent job 2. Open-ended job 3. Self-employment and other, e.g. temporary work
B1	1. The university degree that you hold is specifically required 2. A graduate from a different major could obtain similar results 3. A university degree is not necessary; a high school degree could suffice 4. A qualification lower than high school could suffice
B2	0 (not at all) to 9 (a lot)
B3	(1 = yes, 0 = no)
B4	(1 = yes, 0 = no)
B5	(1 = yes, 0 = no)
B6	1. Not at all 2. Not much 3. Quite 4. Very much
C1	1. The residence province 2. The residence region 3. Abroad or in an Italian region (different from the residence region)
C2	1-(weekly working hours normalized between 0 and 1)
js1	0 (not at all) to 9 (a lot)
js2	0 (not at all) to 9 (a lot)
js3	0 (not at all) to 9 (a lot)
js4	0 (not at all) to 9 (a lot)
js5	0 (not at all) to 9 (a lot)
js6	0 (not at all) to 9 (a lot)
js7	0 (not at all) to 9 (a lot)
js8	0 (not at all) to 9 (a lot)
js9	0 (not at all) to 9 (a lot)
js10	0 (not at all) to 9 (a lot)
js11	0 (not at all) to 9 (a lot)

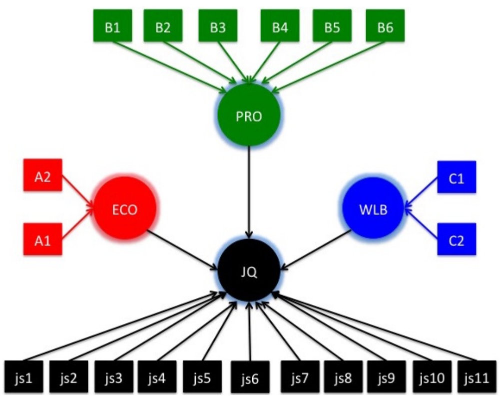

Figure 7: Path diagram with the formative model

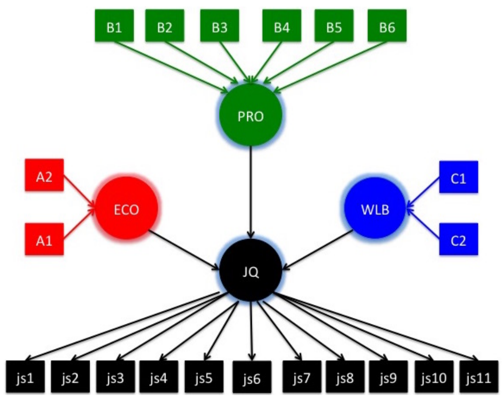

Figure 8: Path diagram with the hybrid model

Table 8: Unidimensionality measures of the hybrid model

Type	C.alpha	DG.rho	Eig.1st
Job Quality	0.783	0.836	3.76

The outer and inner estimations of the path model obtained with the formative and hybrid approaches are shown in Tables 9 and 10, respectively.

Table 9: Estimates of the formative and hybrid measurement models

	Formative model		Hybrid model	
MV	Loading	Communality	Loading	Communality
		Economic		
A1	0.048	0.002	0.424	0.280
A2	1.000	0.999	0.941	0.886
		Professional		
B1	0.371	0.137	0.479	0.229
B2	0.590	0.349	0.620	0.385
B3	0.435	0.189	0.314	0.099
B4	0.724	0.524	0.531	0.282
B5	0.232	0.054	0.142	0.020
B6	0.620	0.384	0.850	0.723
		Work Life Balance		
C1	-0.294	0.087	-0.321	0.103
C2	0.985	0.970	0.980	0.960
		Job Quality		
js1	0.792	0.627	0.521	0.272
js2	0.426	0.182	0.727	0.528
js3	0.441	0.194	0.728	0.530
js4	0.353	0.124	0.563	0.317
js5	-0.236	0.056	0.074	0.006
js6	0.268	0.072	0.511	0.261
js7	0.099	0.010	0.286	0.081
js8	-0.354	0.125	-0.081	0.006
js9	-0.061	0.004	0.085	0.007
js10	0.501	0.252	0.784	0.609
js11	0.549	0.301	0.809	0.655

Table 10: Estimates of the formative and hybrid structural models

	Job Quality - Formative model			Job Quality - Hybrid model		
LV	Estimate	Std. Error	t-value	Estimate	Std. Error	t-value
Economic	0.312	0.017	1.860	0.133	0.016	7.870
Professional	0.311	0.017	1.820	0.508	0.017	2.970
Work Life	0.309	0.017	1.830	0.211	0.017	1.250

In Table 10 we can observe that the different results emerge in the two approaches: in the formative model, all the path coefficients are equal to 0.31, while in the hybrid model the *Professional* dimension presents the biggest coefficient (0.51), whereas the *Economic* and *Work Life Balance* dimensions present a path coefficient estimate equal to 0.13 and 0.21, respectively. In the examination of the overall fit

measures, we have obtained $R^2 = 0.418$ and $GoF = 0.335$ in the formative approach, $R^2 = 0.392$ and $GoF = 0.365$ in the hybrid approach.
Furthermore, we can note that the *Economic* dimension, especially in the formative model, is somewhat strange. It seems that ECO construct is represented by A2 (*Contractual stability*) only, with an evident collinearity (loading equal to 1).
In the next subsection, we present an analysis of the latent score distributions, trying to understand the weakness of the model and to propose an alternative one. In particular, we are interested to evaluate the coefficients of the three dimensions (inner path coefficients), because they represent the weight of each dimension for the general concept *Job Quality*. However, the "right" weights do not exist, then we cannot establish what is the right choice. Anyway, we can proceed with the comparison of these results with results obtained using different methods. In this case, the obtained inner estimates will be compared with the weights obtained by Boccuzzo and Gianecchini (2015) approach.

5.4 Latent scores analysis and adjusted model

The details of the latent scores obtained in both formative and hybrid approaches are shown in Table 11 (descriptive statistics) and in Figures 9 and 10 (distributions plots).

Table 11: Descriptive analysis of the latent score distributions obtained with the formative and hybrid approaches

Formative model approach				
	Economic	Professional	Work Life B.	Job Quality
Min	0.0000	0.0000	0.0000	0.0000
Lower quartile	0.0669	0.5259	0.3768	0.5134
Median	0.0811	0.5962	0.4199	0.6563
Mean	0.5130	0.6266	0.3943	0.6286
Upper quartile	0.9796	0.7656	0.4422	0.7570
Max	1.0000	1.0000	1.0000	1.0000
Hybrid model approach				
	Economic	Professional	Work Life B.	Job Quality
Min	0.0000	0.0000	0.0000	0.0000
Lower quartile	0.2254	0.5990	0.3690	0.5585
Median	0.4693	0.6819	0.4111	0.6723
Mean	0.3880	0.6699	0.3930	0.6557
Upper quartile	0.5597	0.7908	0.4385	0.7706
Max	1.0000	1.0000	1.0000	1.0000

We can note that the ECO score distribution presents an irregular structure, showing a frequency accumulation around to the extreme values. This is due to the collinearity between the construct and the indicator A2 which is dichotomous and to the null loading of the indicator A1. We hypothesize that the causes of this problem are fundamentally two: (*i*) small number of indicators of ECO construct (two) and (*ii*) the high correlation (0.48) between A2-*Contractual stability* (first order indicator) and js1-*Stability and safety of the job* (second order indicator). Figure 11 represents the correlation plot of the job quality data.

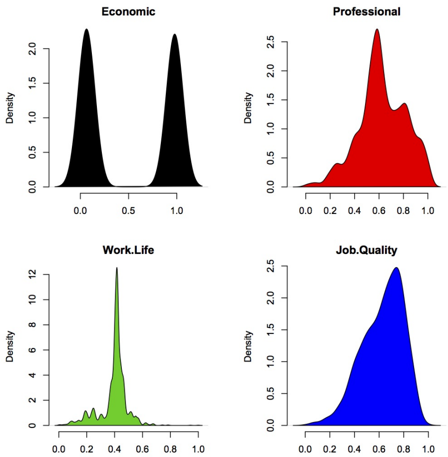

Figure 9: Latent score distributions obtained in the formative approach

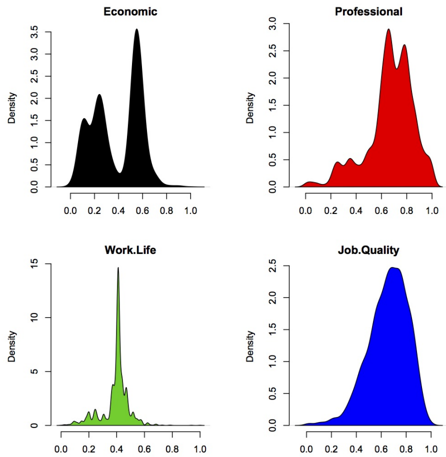

Figure 10: Latent score distributions obtained in the hybrid approach

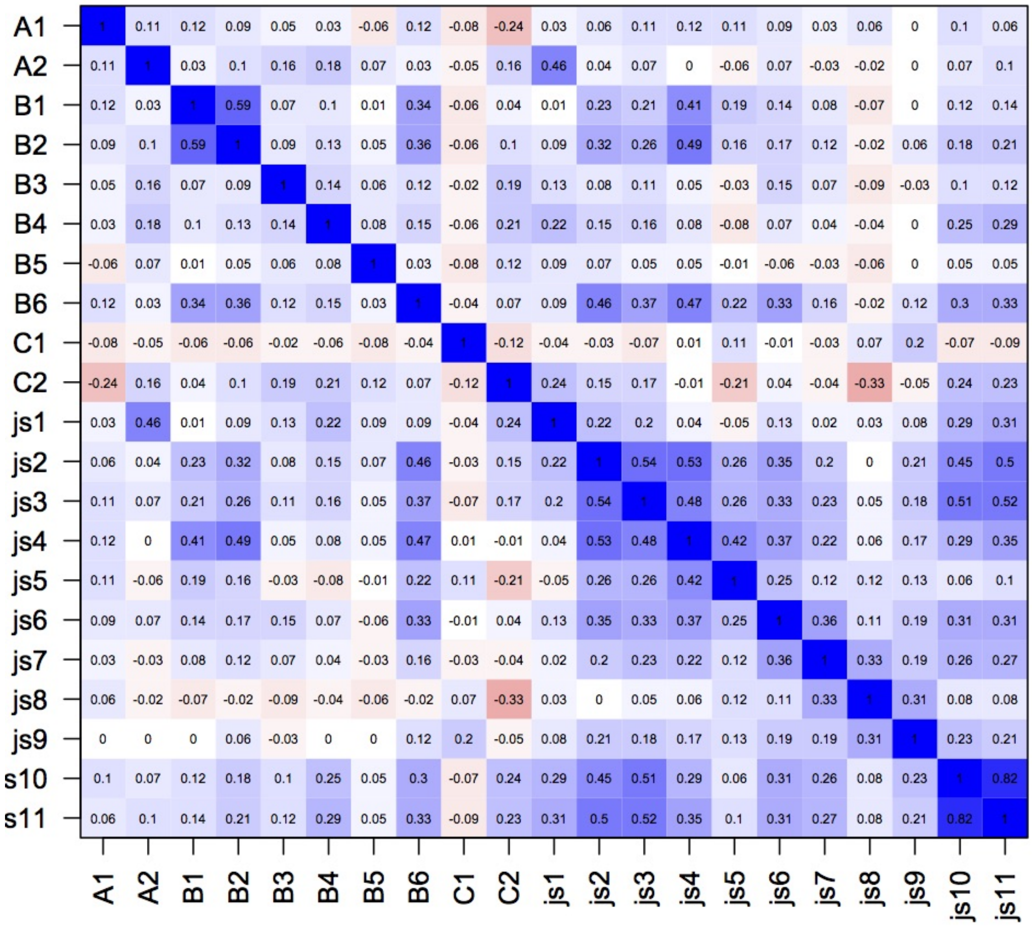

Figure 11: Correlations plot of the Job Quality data

Given the structure of data, the polychoric correlation has been applied. For this reasons, we propose an adjusted version of the model, that we define *Second model*, where the second order indicator js1-*Stability and safety of the job* is not included. In Tables 12 and 13 are shown the outer and inner model estimation.

Table 12: *Second model*: estimates of the measurement models

	Formative model		Hybrid model	
MV	Loading	Communality	Loading	Communality
		Economic		
A1	0.663	0.439	0.813	0.662
A2	0.822	0.676	0.676	0.458
		Professional		
B1	0.569	0.323	0.536	0.288
B2	0.655	0.428	0.637	0.406
B3	0.297	0.088	0.259	0.067
B4	0.445	0.198	0.417	0.174
B5	0.131	0.017	0.098	0.010
B6	0.873	0.761	0.904	0.816
		Work Life Balance		
C1	-0.331	0.110	-0.370	0.137
C2	0.978	0.957	0.968	0.937
		Job Quality		
js2	0.675	0.456	0.766	0.587
js3	0.673	0.453	0.768	0.590
js4	0.618	0.381	0.659	0.434
js5	0.077	0.006	0.293	0.086
js6	0.454	0.206	0.556	0.309
js7	0.190	0.036	0.341	0.116
js8	-0.371	0.138	-0.043	0.002
js9	-0.048	0.002	0.224	0.050
js10	0.712	0.507	0.781	0.609
js11	0.740	0.547	0.805	0.648

Table 13: *Second model*: estimates of the structural models

	Job Quality - Formative model			Job Quality - Hybrid model		
LV	Estimate	Std. Error	t-value	Estimate	Std. Error	t-value
Economic	0.061	0.017	3.670	0.078	0.017	4.600
Professional	0.524	0.017	3.100	0.554	0.017	3.220
Work Life	0.254	0.017	1.520	0.162	0.017	9.540

In this case, the global results do not present particular differences between formative and hybrid approaches. In detail, the results of the outer ECO estimation (Table 12) are surely more stable than the first model and the loadings of A1 and A2 are more "realistic". Conversely, the ECO path coefficients (Table 13) are very low (0.06 in formative approach and 0.08 in hybrid approach). Then, with this model, we solve the problem of the measurement model, but there is a problem in the structural model. The overall fit indices are slightly improved, with $R^2 = 0.4$ and $GoF = 0.367$ in formative approach, $R^2 = 0.379$ and $GoF = 0.374$ in the hybrid

approach.
The details of the latent constructs scores obtained by the *Second model* in both formative and hybrid approaches are shown in Table 14 (descriptive statistics) and in Figures 12 and 13 (distributions plots).

Table 14: Descriptive statistics of the latent scores obtained by the *Second model*

Formative model approach				
	Economic	Professional	Work Life B.	Job Quality
Min	0.0000	0.0000	0.0000	0.0000
Lower quartile	0.3143	0.6357	0.3697	0.5359
Median	0.4041	0.7226	0.4119	0.6428
Mean	0.3769	0.6966	0.3935	0.6254
Upper quartile	0.4581	0.8155	0.4446	0.7245
Max	1.0000	1.0000	1.0000	1.0000
Hybrid model approach				
	Economic	Professional	Work Life B.	Job Quality
Min	0.0000	0.0000	0.0000	0.0000
Lower quartile	0.2954	0.6429	0.3626	0.5475
Median	0.3487	0.7235	0.4040	0.6607
Mean	0.3432	0.6955	0.3892	0.6413
Upper quartile	0.4135	0.8164	0.4382	0.7506
Max	1.0000	1.0000	1.0000	1.0000

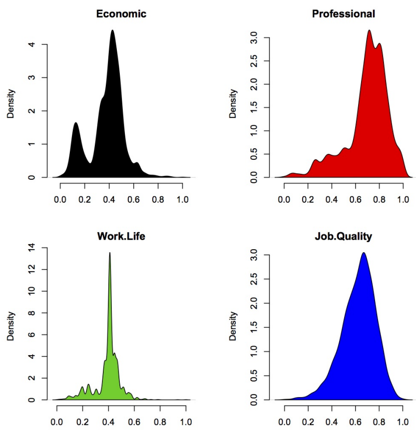

Figure 12: Latent score distributions obtained by the *Second model*: Formative approach

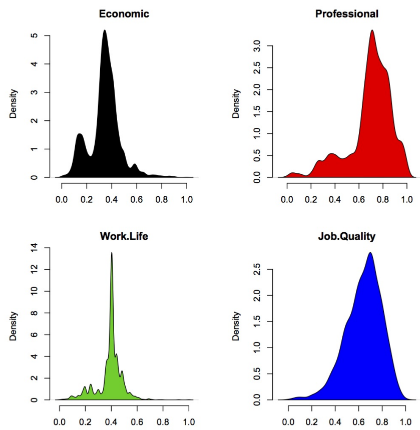

Figure 13: Latent score distributions obtained by the *Second model*: Hybrid approach

The latent score distributions present an improvement of the *Economic* construct. In fact the plot and the descriptive statistics show a better structure of data, very similar in both cases (formative and hybrid).
However, we need to solve the problem of the structural model. We believe that this problem could derive from the exclusion of the unique second order indicator associated to the *Economic* construct (i.e. js1).
The alternative model, that we define *Third model*, is represented by the path di-

agrams in Figures 14 and 15. Then, given the high correlation of js1 with A2, we have provided to include the second order indicator js1 in the *Economic* construct.

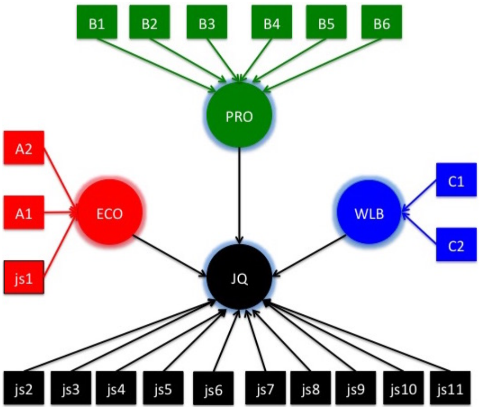

Figure 14: Path diagram of the *Third model* - Formative model

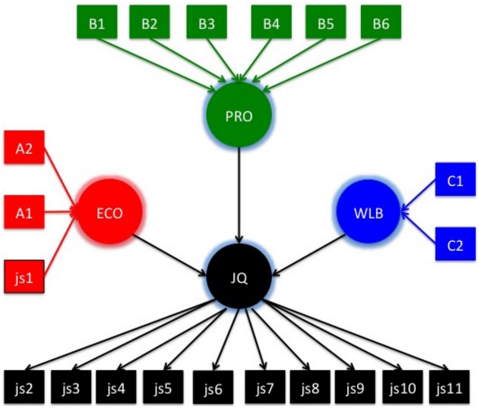

Figure 15: Path diagram of the *Third model* - Hybrid model

The outer and inner estimations of the path model obtained with the formative and hybrid approaches are shown in Tables 15 and 16, respectively.

Table 15: *Third model*: estimates of the measurement models

	Formative model		Hybrid model	
MV	Loading	Communality	Loading	Communality
		Economic		
A1	0.188	0.035	0.341	0.116
A2	0.345	0.118	0.319	0.101
js1	0.980	0.960	0.945	0.894
		Professional		
B1	0.493	0.243	0.513	0.264
B2	0.614	0.377	0.625	0.391
B3	0.305	0.093	0.262	0.070
B4	0.548	0.300	0.449	0.201
B5	0.141	0.020	0.106	0.011
B6	0.843	0.711	0.898	0.806
		Work Life Balance		
C1	-0.338	0.114	-0.364	0.133
C2	0.977	0.954	0.970	0.940
		Job Quality		
js2	0.675	0.455	0.761	0.579
js3	0.654	0.428	0.764	0.584
js4	0.505	0.255	0.618	0.381
js5	-0.164	0.027	0.100	0.010
js6	0.399	0.159	0.535	0.287
js7	0.169	0.029	0.341	0.117
js8	-0.333	0.111	-0.038	0.001
js9	0.066	0.004	0.257	0.066
js10	0.773	0.597	0.809	0.655
js11	0.801	0.642	0.833	0.693

Table 16: *Third model*: estimates of the structural models

	Job Quality - Formative model			Job Quality - Hybrid model		
LV	Estimate	Std. Error	t-value	Estimate	Std. Error	t-value
Economic	0.178	0.017	1.060	0.198	0.017	1.180
Professional	0.452	0.017	2.680	0.519	0.017	3.080
Work Life	0.271	0.017	1.610	0.145	0.017	8.640

Observing the tables, it seems that the problem of extreme reduction of the ECO weight has been solved in both approaches (0.18 in formative approach and 0.20 in hybrid approach). Furthermore, with respect to the previous models, also the measurement models seem improved, especially in *Economic* dimension. The overall fit indices are almost unvaried, with $R^2 = 0.408$ and $GoF = 0.359$ in formative approach, $R^2 = 0.403$ and $GoF = 0.374$ in the hybrid approach. The details of the latent constructs scores obtained by the *Third model* in both formative and hybrid approaches, are shown in Table 17 (descriptive statistics) and Figures 16 and 17 (distribution plots).

Table 17: Descriptive statistics of the latent scores obtained by the *Third model*

Formative model approach				
	Economic	Professional	Work Life B.	Job Quality
Min	0.0000	0.0000	0.0000	0.0000
Lower quartile	0.6649	0.6335	0.3708	0.5718
Median	0.7545	0.7021	0.4132	0.6846
Mean	0.6802	0.6954	0.3934	0.6671
Upper quartile	0.7966	0.8244	0.4483	0.7786
Max	1.0000	1.0000	1.0000	1.0000
Hybrid model approach				
	Economic	Professional	Work Life B.	Job Quality
Min	0.0000	0.0000	0.0000	0.0000
Lower quartile	0.5314	0.6433	0.3634	0.5729
Median	0.5918	0.7190	0.4049	0.6898
Mean	0.5669	0.6967	0.3897	0.6693
Upper quartile	0.6863	0.8161	0.4385	0.7816
Max	1.0000	1.0000	1.0000	1.0000

The details of the latent scores show that there are not particular differences between formative and hybrid approaches, except for the *Economic* construct. In the next subsection we present a comparison between the structural model of the *Third model* and the structural model of Boccuzzo and Gianecchini (2015).

5.5 Structural models comparison

In this subsection we limit our analysis to the inner model, and we will compare the results with the weights obtained by Boccuzzo and Gianecchini (2015). In their paper, the authors use a traditional approach for the construction of the composite indicator, and weights of the dimensions were calculated on the basis of indirect opinions of a sample of graduates, i.e. stated-preference approach (Decancq and Lugo, 2013). In the same article, weights were validated with a hedonic approach.

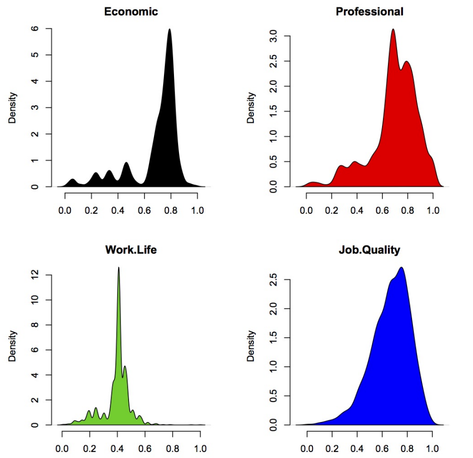

Figure 16: Latent score distributions obtained by the *Third model*: Formative approach

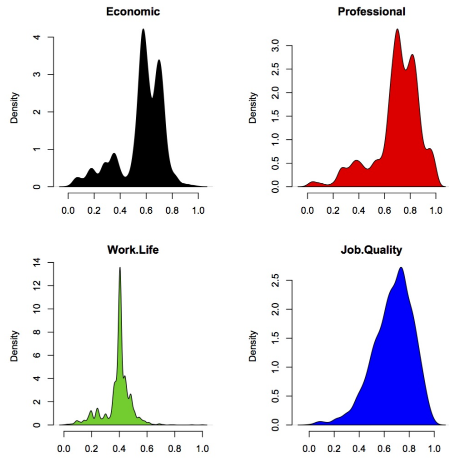

Figure 17: Latent score distributions obtained by *Third model*: Hybrid approach

These approaches are completely different and we cannot affirm that they represent the "golden standard". Anyway, a comparison among approaches could be useful in order to better understand the results.
In Table 18 we compare the weights (normalized to sum up to 1) of the formative and hybrid models to the mean of the weights of Boccuzzo and Gianecchini (2015). The sample is exactly the same.

Table 18: Structural models comparison: Model 1, 2, 3 and B & G

LVs	Model 1		Model 2		Model 3		Mean
	Formative	Hybrid	Formative	Hybrid	Formative	Hybrid	B & G
ECO	0.338	0.156	0.073	0.098	0.198	0.230	0.223
PRO	0.332	0.596	0.625	0.698	0.502	0.602	0.621
WLB	0.330	0.248	0.303	0.204	0.301	0.168	0.157

We can note that, with respect to the three proposed models, the results most similar to B & G are obtained by the *Third model*. Furthermore, with respect to the formative approach, the hybrid model presents results very similar to those obtained by Boccuzzo and Gianecchini (2015).

6 Conclusions

In this work we have tried to build a composite indicator through the methodology of the path models. The path model structure, differently to composite indicators, provides a measurement models of second order that establishes the relationship between the overall latent construct (CI) and the external observed variables (second order indicators). This second order relationships, with respect to first, can be (*i*) formative, if the overall latent construct is supposed to be formed by its indicators; (*ii*) reflective, if the observed variables are considered as being caused by the overall

latent construct (Sanchez, 2013).
We have testing the path models approach for CI construction through the application of PLS method (Tenenhaus et al., 2005; Vinzi et al., 2010b; Monecke and Leisch, 2012) on toy data and real data. The application on real data concerns on the measurement of job quality. In this case, we have built formative models (with formative second order indicator) and hybrid models (with reflective second order indicators), obtaining different results. Through a comparison with the recent JQCI application of Boccuzzo and Gianecchini (2015), the evaluation model criteria suggest that the hybrid approach is preferable to the formative approach.
However, several aspects should be considered in the construction of a CI through the PLS-PM approach:

1. The number of elementary indicators for each dimension should be not too low; almost three indicators are necessary.
   Furthermore, if the indicators are not quantitative, especially dichotomous, this recommendation become even more important, because the scores distribution of the dimension could be too much influenced by the dichotomous observed variables.

2. The correlation among the elementary indicators should be carefully analyzed, in order to correct allocate the indicators to the dimensions and for the stability of the model. In our case, the satisfaction for the job stability resulted to be an indicator of the ECO dimension and, when we have moved it from the JQ measurement to the ECO measurement, the final model has improved.

3. The hybrid model, based on reflective measurement of the general latent concept (second order LV), is less influenced by the correlation among elementary indicators (Table 9, estimates related to A1), given that the estimation procedure is based on simple regression, not multiple regression as in formative approach.

7 Appendix A: unidimensionality measures

Cronbach's alpha.
The Cronbach's alpha is a coefficient that is intended to evaluate how well a block of indicators measure their corresponding latent construct (Bland et al., 1997; Vinzi et al., 2010b). This index is an average inter-correlation between indicators of a reflective latent construct. If a block of manifest variables is unidimensional, they have to be highly correlated, and consequently we expect them to have a high average inter-correlation. Among several alternative and equivalent formulas, this index for the p-th latent construct can be expressed as:

$$\alpha_p = \frac{\sum_{j \neq j^*=1}^{J_p} cor(x_{jp}, x_{j^*p})}{J_p + \sum_{j \neq j^*=1}^{J_p} cor(x_{jp}, x_{j^*p})} \cdot \frac{J_p}{J_p - 1},$$

where x_{jp} is the j-th manifest variable of the p-th reflective latent construct, x_{j^*p} is another j-th manifest variable of the p-th reflective latent construct and J_p is the number of the manifest variable of the p-th reflective latent construct. Note that the computation of the Cronbach's alpha requires the observed variables to be standardized.

Dillon-Goldstein's rho.
Dillon-Goldstein's rho is a Cronbach's alpha generalization. This index measures the unidimensionality of a LV through the correlations between the reflective latent construct and each manifest variable of the corresponding block, i.e. the loadings (Vinzi et al., 2010b; Sanchez, 2013). Formally, the index for the p-th latent construct is:

$$\rho_p = \frac{\left(\sum_{j=1}^{Jp} \lambda_{jp}\right)^2}{\left(\sum_{j=1}^{Jp} \lambda_{jp}\right)^2 + \sum_{j=1}^{Jp}\left(1-\lambda_{jp}^2\right)},$$

where λ_{jp} is the loading between the j-th manifest variable and the p-th latent construct.

8 Appendix B: local and global fit measures

Communality.
Communality is an index of local fit calculated on each manifest variable and the latent construct. This index is calculated with the purpose to check that indicators in a block are well explained by its latent variable. Communalities are simply squared loadings and they measure the part of the covariance between a latent variable and its indicator that is common to both (Sanchez, 2013). Then, the communality for the j-th manifest variable of the p-th latent construct is calculated as:

$$com(x_{jp}, y_p) = cor(x_{jp}.y_p)^2 = \lambda_{jp}^2,$$

where x_{jp} is the j-th manifest variable of the p-th latent construct and y_p is the p-th latent construct.

Coefficient of determination R^2.
For each regression in the structural model we have an R^2 that is interpreted similarly as in any multiple regression analysis. R^2 indicates the amount of variance in the endogenous latent variable explained by its independent latent variables (Vinzi et al., 2010b; Sanchez, 2013). Values for the R-squared can be classified in three categories:

1. *low* if $R^2 < 0.20$
2. *moderate* if $0.20 < R^2 < 0.50$
3. *high* if $R^2 > 0.50$

Goodness of fit GoF.
The GoF index is a pseudo Goodness of fit measure that accounts for the model

quality at both the measurement and the structural models. *GoF* is calculated as the geometric mean of the average communality and the average R^2 value. Formally,

$$\sqrt{\overline{com} \cdot \overline{R^2}}.$$

This index is more applicable to reflective indicators than to formative indicators (Vinzi et al., 2010b; Russolillo, 2012; Sanchez, 2013). However, you can also use the *GoF* index in presence of formative blocks, in which case more importance will be given to the average R^2. Unfortunately, there is also no guidance about what number could be considered a good *GoF* value. However, *GoF* can be seen as an index of average prediction for the entire model. For instance, a *GoF* value of 0.78 could be interpreted as if the prediction power of the model is of 78%.

References

Anand, S. and Sen, A. (1994). Human development index: methodology and measurement. Technical report, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Bagozzi, R. P. and Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40(1):8–34.

Becker, J. M., Klein, K., and Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: guidelines for using reflective-formative type models. *Long Range Planning*, 45(5):359–394.

Bland, J. M., Altman, D. G., et al. (1997). Statistics notes: Cronbach's alpha. *Bmj*, 314(7080):572.

Boccuzzo, G. and Gianecchini, M. (2015). Measuring young graduates? job quality through a composite indicator. *Social Indicators Research*, 122(2):453–478.

Bollen, K. A. (1998). *Structural equation models*. Wiley Online Library.

Bollen, K. A. and Ting, K. f. (2000). A tetrad test for causal indicators. *Psychological methods*, 5(1):3.

Cenfetelli, R. T. and Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *Mis Quarterly*, pages 689–707.

Decancq, K. and Lugo, M. A. (2013). Weights in multidimensional indices of well-being: An overview. *Econometric Reviews*, 32(1):7–34.

Diamantopoulos, A. (1999). Viewpoint-export performance measurement: Reflective versus formative indicators. *International Marketing Review*, 16(6):444–457.

Diamantopoulos, A. and Winklhofer, H. M. (2001). Index construction with formative indicators: an alternative to scale development. *Journal of marketing research*, 38(2):269–277.

Edwards, J. R. (2010). The fallacy of formative measurement. *Organizational Research Methods*.

Edwards, J. R. and Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological methods*, 5(2):155.

Fabbris, L. (2012). Preface: Graduates as informants of higher education effectiveness. *in Fabbris, L. (Ed.), Indicators of Higher Education Effectiveness*, pages 5–11.

Hair, J. F., Ringle, C. M., and Sarstedt, M. (2011). Pls-sem: Indeed a silver bullet. *The Journal of Marketing Theory and Practice*, 19(2):139–152.

Hauser, R. M. and Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. *Sociological methodology*, 3:81–117.

Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics*, 25(1):107–120.

Henseler, J., Ringle, C. M., and Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in international marketing*, 20:277–319.

Hsu, S.-H., Chen, W.-H., and Hsieh, M.-J. (2006). Robustness testing of PLS, LISREL, EQS and ANN-based SEM for measuring customer satisfaction. *Total Quality Management & Business Excellence*, 17(3):355–372.

James, L. R. and Jones, A. P. (1980). Perceived job characteristics and job satisfaction: An examination of reciprocal causation. *Personnel Psychology*, 33(1):97–135.

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of consumer research*, 30(2):199–218.

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8 user's reference guide*. Scientific Software International.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford press.

Law, K. S. and Wong, C.-S. (1999). Multidimensional constructs m structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management*, 25(2):143–160.

MacKenzie, S. B., Podsakoff, P. M., and Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, 90(4):710.

Monecke, A. and Leisch, F. (2012). semPLS: structural equation modeling using partial least squares. *Journal of Statistical Software*, 48(3):1–32.

OECD (2008). Handbook on constructing composite indicators. *Methodology and User Guide.*

Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., and Lee, J. Y. (2003). The mismeasure of man (agement) and its implications for leadership research. *The Leadership Quarterly*, 14(6):615–656.

Ringle, C. M., Götz, O., Wetzels, M., and Wilson, B. (2009). On the use of formative measurement specifications in structural equation modeling: A Monte Carlo simulation study to compare covariance-based and partial least squares model estimation methodologies. *METEOR Research Memoranda (RM/09/014).*

Russolillo, G. (2012). Non-metric partial least squares. *Electronic Journal of Statistics*, 6:1641–1669.

Sanchez, G. (2013). Pls path modeling with r. *Online, January.*

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., and Lauro, C. (2005). Pls path modeling. *Computational statistics & data analysis*, 48(1):159–205.

Vinzi, V., Chin, W. W., Henseler, J., and Wang, H. (2010a). Handbook of partial least squares: Concepts, methods and applications. *Radboud University Nijmegen.*

Vinzi, V. E., Trinchera, L., and Amato, S. (2010b). Pls path modeling: from foundations to recent developments and open issues for model assessment and improvement. In *Handbook of partial least squares*, pages 47–82. Springer.

Wetzels, M., Odekerken-Schroder, G., and Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. *Management Information Systems Quarterly*, 33(1):11.

Wilcox, J. B., Howell, R. D., and Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, 61(12):1219–1228.