

FROM IMPLICIT ASSUMPTIONS TO EXPLICIT TESTING: MEASURING CONSTRUCTS WITHIN A MULTILEVEL FRAMEWORK AND EVALUATING THEIR PSYCHOMETRIC ISOMORPHISM

SEMIRA TAGLIABUE
CATHOLIC UNIVERSITY OF BRESCIA

ANGELA SORGENTE
MARGHERITA LANZ
CATHOLIC UNIVERSITY OF MILANO

Researchers who use a multilevel framework often aggregate scores measured at Level 1 in order to measure the construct at Level 2. In doing so, they implicitly assume that the construct is measured at both the individual and cluster level, that the ontology of the construct is the same at both levels, and that reliability is good at both levels. The aim of this paper is to present a seven-step approach that allows one to explicitly test those assumptions. Multilevel confirmatory factor analysis, a statistical technique that control the nonindependence of data in measurement models testing a factor structure at the within-level and between-level simultaneously, provides an analytic framework in which it is possible to test psychometric isomorphism and reliability. Recommendations about what to do when assumptions are verified or not are provided in the Section “Discussion.” The Appendix reports Mplus syntaxes necessary to run each step of the analysis.

Keywords: Multilevel confirmatory factor analysis; Psychometric isomorphism; Level-specific reliability; Nested data; Measurement assumptions.

Correspondence concerning this article should be addressed to Semira Tagliabue, Department of Psychology, Catholic University of Brescia, Via Trieste 17, 25121 Brescia (BS), Italy. Email: semira.tagliabue@unicatt.it

In social science research, data are often hierarchically structured (i.e., micro-level units are nested with a macro-level). For example, in educational research, pupils are nested within schools; in organizational research, employees are nested within companies; in family research, members are nested within families; in longitudinal studies, time points measures are nested within subjects; and so on. In recent decades, researchers have become more aware that nested data have to be analyzed taking into consideration the dependencies between observations within groups, in order to avoid biased results (Geiser, 2013). This awareness has increased the utility of multilevel models (Geldhof, Preacher, & Zyphur, 2014).

Multilevel models are often used in research that aims to test the relationship between variables, taking into consideration the different levels which characterize the hierarchical structure. For example, in organizational research, an important issue is to analyze predictors of employees' experience, which could be determined by employees' characteristics (Level 1; for instance, seniority within the company) as well as contextual elements of the teams and companies they belong to (higher levels; for instance, company size) (Liu, Chen, & Yao, 2011).¹ In multilevel models, variance of the dependent variable (employees' experience measured at Level 1) is explained by both predictors at Levels 1 and 2 and cross-level interactions. Regarding predictors at Level 2, researchers can choose predictors directly measured at that level (for instance, size of

the company) or predictors measured at Level 1 but aggregated at Level 2 in order to measure a cluster's property (for instance average employee's seniority, a Level 2 predictor).

Although some literature presents solutions for how to do the aggregation (Dixon & Cunningham, 2006; Suzuki, Yamamoto, Takao, Kawachi, & Subramanian, 2012), very often, aggregation is simply done by averaging the scores of members of each cluster, in order to obtain a cluster level variable. For instance, just as a few recent examples, Alm, Låftman, Sandahl, and Modin (2019) measured school leadership, school ethos, and teacher cooperation and consensus by averaging the corresponding scores provided by teachers within each school; Thomas et al. (2017) measured collective identity motives averaging, within each sport team, athletes' score on the Social Identity Motives Scale. When researchers average scores at Level 1 in order to create a variable at Level 2, or cluster level, they implicitly assume some measurement properties of the instrument. Those implicit assumptions are: 1) it is possible to measure the construct at both the individual (or lower) and cluster level; 2) the construct at the cluster level has the same dimensions as the construct at the individual (or lower) level; 3) the construct at the cluster level has the same meaning as the construct at the individual (or lower) level; 4) the reliability is the same at the cluster and individual (or lower) levels. Those assumptions are implicit because, as said, in many papers very little reasoning is provided about the psychometric characteristics of instruments used in multilevel models. The present paper aims to stimulate researchers' awareness about those implicit assumptions, and to show how to explicitly test them in order to improve understanding of constructs and their interpretation of multilevel regression models.

Nonindependent data produce a multilevel or nested structure that also needs to be taken into account in the psychometric analysis of the data. Currently, techniques and statistical software are available to analyze multilevel data, but these techniques have been mostly used to analyze relationships between constructs at different levels of analysis rather than the psychometric properties of measurement models (Dedrick & Greenbaum, 2011).

FIRST IMPLICIT ASSUMPTION: IT IS POSSIBLE TO MEASURE THE CONSTRUCT AT BOTH THE INDIVIDUAL AND CLUSTER LEVELS

The first assumption is based on the fact that, because there is a nested structure of data (e.g., employees within companies), a specific construct can be different both within clusters and between them. So, for instance, employees within the same company are different in their evaluation of their experience (within variance), and companies are different in how much, on average, their employees are experts (between variance). However, the implicit assumption should be tested before using variables at both levels. Indeed, there is a need to verify whether the between variance is present, that is, whether individual ratings are affected by the fact that they belong to different companies. If not, it makes no sense to aggregate the scores at the company level because companies are not different. Usually, to test that assumption, the intraclass correlation coefficient (ICC) is computed. The ICC measures the proportion of variance in the variable that is explained by the grouping structure of the hierarchical model. It is important to specify that, when the construct of interest (e.g., experience) is measured with an instrument composed of a certain number of items, there is a need to test that assumption for each item; indeed, we want to be sure that belonging to a specific cluster partially explains how each employee evaluates each item. Each item's ICC represents the ratio between how much the intercepts (means) of the item varies between groups and the total variation (sum of the intercepts' variation between groups and the variation within groups). When groups' intercepts differ but within any group there is no variation, ICC will equal 1; on the contrary, when groups' intercepts are all the same and

there is within-group variation, ICC approaches 0 and hierarchical modeling is not appropriate (Garson, 2013). An ICC equal or higher than .05 indicates the need to consider the nonindependence of data (Dyer, Hanges, & Hall, 2005). Some authors (e.g., Peugh, 2010) also suggest to accompany the ICC with the estimation of the design effect (DE), as it corrects the ICC estimation for the sample design and cluster size (e.g., how many members there are, on average, in each company included in data collection). In particular, $DE = 1 + ICC(c-1)$, where c is the average group size. Design effect's values higher than 2 are usually obtained when the nonindependence of data cannot be ignored (Satorra & Muthen, 1995).

SECOND IMPLICIT ASSUMPTION: THE CONSTRUCT AT THE CLUSTER LEVEL HAS THE SAME DIMENSIONS OF THE CONSTRUCT AT THE INDIVIDUAL LEVEL (CONFIGURAL ISOMORPHISM)

After having verified the first assumption, the researcher could think, implicitly, that the structure of the construct (number of factors, items measuring each factor) is the same at the individual (i.e., employees) and cluster (i.e., companies) levels. However, it is possible that is not the case. For instance, Margola, Fenaroli, Sorgente, Lanz, and Costa (2019) showed that the Family Relationships Index (FRI) — an instrument measuring the quality of family relationships — can have three dimensions at the individual level (conflict, cohesion, communication) and just one dimension at the family level (relationship quality). So, before adopting the implicit assumption that the within and the between level construct have the same structure, researchers should test this assumption.

Testing this assumption means testing a measurement model adopting a multilevel framework, multilevel confirmatory factor analysis (MCFA), in order to simultaneously test a factor structure at the within-level (e.g., experience score related to the single employee) and at the between-level (e.g., average employees experience score related to the specific company), identifying individual-level and group-level (or cluster-level) latent factor(s) that explain the variance of items' scores measured at individual-level. MCFA allows taking into account the presence of two or more levels in relation to the covariance matrix and the factorial structure (Dedrick & Greenbaum, 2011). With regard to the covariance matrix, MCFA decomposes the total sample covariance matrix (S_T) into the sample pooled within-group covariance matrix (S_{PW}) and the sample between-group covariance matrix (S_B). The S_T is an estimator of the total population matrix, Σ_T — with $\Sigma_T = \Sigma_W + \Sigma_B$, where Σ_W and Σ_B represent respectively the within- and between-group population matrices. While in classic CFA, analyses are performed on the whole sample covariance matrix S_T , in the multilevel framework, interest focuses primarily on the two decomposed sample matrices: S_{PW} and S_B , where the analysis of S_{PW} , which is a consistent and unbiased estimator of Σ_W , is a within matrix representing within-level variation, and S_B , which is a consistent and unbiased estimator of $\Sigma_W + c\Sigma_B$ (where c represents the group/class size), is a between matrix reflecting across-group variation (Li, Duncan, Harmer, Acock, & Stoolmiller, 1998). In other words, using these two matrices, it is possible to separately model the covariance among items that is due to individuals and the one that is due to groups.

With regard to the factor structure, MCFA allows for testing the structure of an instrument separately at the two levels. Specifically, the two within- and between-level matrices are separately and simultaneously used to analyze the factor structure at each level. So, not only does MCFA analyze the covariances of the two matrices separately, but it also allows for different structures at each level. This is a relevant issue, as some constructs may have different factor structure at different levels (lack of configural isomorphism). When the factorial structure is the same at the two levels, configural isomorphism is found. Indeed, configural

isomorphism means that the factor structures of lower- and higher-level constructs are equivalent. Specifically, the same number of factors and the pattern of zero and nonzero factor loadings are expected to hold across levels. The lack of configural isomorphism indicates that the cluster-level construct is different (has a different structure and is measuring a different concept) than the individual-level construct. This shows that the structure of inter-group variability is fundamentally different from that of inter-individual variability.

THIRD IMPLICIT ASSUMPTION: THE CONSTRUCT AT THE CLUSTER LEVEL HAS THE SAME MEANING AS THE
CONSTRUCT AT THE INDIVIDUAL LEVEL (PSYCHOMETRIC ISOMORPHISM)

Having a good fit when the same factorial model (e.g., three factors) is required at both levels is not sufficient to affirm that the meaning of the construct is the same across levels. Thus, the researcher should be aware of this third possible implicit assumption. It is possible to affirm that the meaning of latent constructs is the same only when items' factor loadings are also equivalent across levels. Tay, Woo, and Vermunt (2014) proposed the expression "psychometric isomorphism"² to indicate the condition in which both the factorial structure (configural isomorphism) and the factor loadings (metric isomorphism) are the same across levels. Metric isomorphism exists if factor loadings are equivalent across levels. The presence of both configural and metric isomorphism would suggest that the interpretation of the common factors is similar across levels. If, instead, a lack of metric isomorphism is found when configural isomorphism is present, it means that there are differences in the defining characteristics of constructs between levels. Specifically, the content of an item is more representative (i.e., higher factor loading) of the meaning of the construct at one level than at the other level. Only if researchers find both configural and metric isomorphism can conclude that they are measuring the same construct at the two levels.

The concept of "psychometric isomorphism" has to be differentiated from the concept of "measurement invariance." In particular, measurement invariance refers to the comparison across different measurement models at the same level; the focus is on the comparison of the measurement models between different groups (e.g., male and female as in Elosua & Hermosilla, 2013), different time points (e.g., Time 1 and Time 2 as in Sorgente, Tagliabue, & Lanz, 2019), different relationship types (e.g., same perceiver's relationships with father, with mother, and with romantic partner; Tagliabue & Lanz, 2014), or different members of the same groups (e.g., Partner 1 and Partner 2 as in Claxton, DeLuca, & VanDulmen, 2015). The general idea is that the different measurement models belong to the same levels (e.g., individual-level for the first, third, and fourth example, time-level for the second example). Instead, when researchers test psychometric isomorphism, the compared measurement models belong to different levels, allowing one to distinguish constructs at the within and between levels (e.g., the employee and the company level).

FOURTH IMPLICIT ASSUMPTION: RELIABILITY IS THE SAME AT THE CLUSTER AND INDIVIDUAL LEVELS
(LEVEL-SPECIFIC RELIABILITY)

Even if researchers are aware of the three previous assumptions, and they test them before aggregating, they could implicitly assume that, having the same meaning at both levels, the reliability is good at both levels. Researchers using nonindependent data should consider that the reliability of a score (i.e., ratio of true score variance over total variance at that level; Jak & Jorgensen, 2017) depends not only on the factor loadings of the items but also on their residual variance. Consequently, the presence of psychometric isomorphism is not a sufficient condition to expect that the scale is equally reliable at both levels. Reliability

needs to be measured separately for each level (Geldhof et al., 2014). This is a relevant issue considering that multilevel constructs, that is phenomena that are potentially meaningful both at the level of individuals and at one or more levels of aggregation (Forer & Zumbo, 2011), as well as instruments adopted to measure them, can work differently at different levels of analysis (Dedrick & Greenbaum, 2011).

According to Geldhof and colleagues (2014), accounting for multilevel variability has been firmly established for hypothesis testing but has been largely ignored in the context of estimating a scale's reliability. Reliability estimation is usually computed at the single level even with nested data, ignoring the option of accounting for multilevel variability. Ignoring hierarchical data structures can bias reliability estimation for a desired level of analysis. Single-level reliability estimates therefore do not necessarily reflect true scale reliability at any single level of analysis.

Geldhof and colleagues (2014) provide formulas to estimate reliability separately for the within- and between-level factors. Specifically, they provided syntaxes for the three more commonly utilized reliability estimates: alpha (α), omega or composite reliability (ω), and maximal reliability (H). In the current paper, we present and utilize only ω . We decided not to present alpha because it is a consistent estimate of reliability only when all items load on a single underlying construct and when all items represent that construct equally well (i.e., essential τ -equivalence), a condition that is rarely assumable. Furthermore, we decided not to present H because, after performing simulation studies, Geldhof and colleagues (2014) stated that "given the propensity for H to overestimate its population value at the between level, we cannot recommend its use in empirical multilevel research" (p. 89).

Omega (ω) has been discussed by several authors (e.g., Bentler, 2009; Raykov, 1997). It is a SEM-based reliability index, conceptually similar to alpha. Zinbarg, Revelle, Yovel, and Li (2005) reported that, even when the assumptions of the essentially τ -equivalent model are met, omega performs at least as well as alpha. However, under violations of τ -equivalence, omega outperforms alpha and is clearly the preferred choice. The formula to compute omega (Geldhof et al., 2014) is the following:

$$\omega = \frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \theta_{ii}}$$

where λ_i represents the factor loading of item i on a single common factor and θ_{ii} represents the unique variance of item i . This formula has to be performed twice, one for each level. Between-level reliability will be based on factor loadings and residual variances of the between-level factor solution, while the within-level reliability will be based on factor loadings and residual variances of the within-level solution. If metric isomorphism holds, the factor loadings will be the same across levels and only the residual variances will differ between the two equations.

AIM

As should be clear, when those four assumptions remain implicit, researchers interpret their multilevel regression models' findings as measuring the same meaning and the same reliable construct(s) at the within and between levels. But, that may not be true, and only explicitly testing each assumption would verify it.

The current paper aims to illustrate how to test the four assumptions in order to increase researchers' awareness related to measurement of psychometric properties. In particular, a multistep procedure that allows testing the measurement model of the construct at both the within and between levels is introduced to the

reader. Specifically, when researchers assume that their constructs have the same meaning at both within and between levels (i.e., this is true any time they aggregate individual score to obtain group-level score), they should test this assumption by running an isomorphic MCFA model. This consists of performing MCFA, psychometric isomorphism, and level-specific reliability analyses. This practice has been systematized into a multistep procedure composed of seven steps, which integrate the procedure proposed by Dyer et al. (2005) with the suggestions proposed by Tay and colleagues (2014) and Geldhof and colleagues (2014). Steps 1 and 2 allow researchers to test the first assumption (it is possible to measure the construct at both individual and cluster level); Steps 3, 4, and 5 are necessary to perform MCFA and to test the second assumption (configural isomorphism), while the Steps 6 and 7 are necessary to test, respectively, the third and fourth assumptions, that is, testing for psychometric isomorphism (as suggested by Tay et al., 2014) and level-specific reliability (as suggested by Geldhof et al., 2014).

The presentation of these seven steps was realized through an illustrative application regarding the multilevel measurement model of the Social Identity Motives Scale of Thomas et al. (2017). That scale has been proposed as a unidimensional scale measuring the social motivations that predispose people toward a specific social identification (Vignoles & Manzi, 2014). The development of that scale has been based on the motivated identity construction theory (MICT; Vignoles, 2011). According to MICT, people are motivated to identify themselves with group (e.g., with the company where they work) in order to positively feel about themselves (self-esteem motive); to feel distinguished from others (distinctiveness motive); to feel included and accepted (belonging motive); to feel that their lives are meaningful (meaning motive); to feel that their past, present, and future are connected (continuity motive); and to feel competent and capable of influencing their environments (efficacy motive). According to MICT (Vignoles, 2011), those six identity motives (self-esteem, distinctiveness, belonging, meaning, continuity, and efficacy) hold to both personal and social level. For example, a self-esteem motive at *personal* level could be “Being a member of this team makes me see myself positively” (i.e., my identification with the company is motivated by the positive esteem I have for myself thanks to my belonging), while a self-esteem motive at *social* level could be “I see this team positively” (i.e., my identification with the company is motivated by the positive esteem I have for the company). The Social Identity Motives Scale of Thomas et al. (2017) measures the *social* motives that motivates people to identify themselves with the group.

Assuming a multiple informant approach (i.e., more informants evaluating the same entity; Lanz, Sorgente, & Tagliabue, 2018), the instrument has here been adapted to measure how employees perceive their company, with the idea that, if employees see their company in a positive way, they have a *social* motive to develop an identification with the company. As stated by Thomas et al. (2017), it is important to analyze this construct within a multilevel framework because in the responses that each individual gives to the items there is both an individual component (i.e., how the single individual perceive the group) and a collective component (i.e., how “group members collectively perceive the group,” Thomas et al., 2017, p. 10). Thomas et al. (2017) adopted this scale to study social identification in sport teams (i.e., athletes nested with teams), while we used it to investigate the social identity motives of Italian employees in different companies. Consequently, we expect that the construct at the individual level is a measure of “how the single employee perceives the company” (a score that allows for comparison of social motives across individuals), while at the group level it is a measure of “how each company is perceived by its members” (a score that allows comparison across companies).

METHOD

Participants

Participants in the present study were drawn from Talenti senza età: donne e uomini over50 e il lavoro [Ageless talents: Over 50 women and men and their job], a study launched in 2017 by Valore D and Università Cattolica del Sacro Cuore to examine which challenges, expectations, obstacles, and opportunities characterize employees of Italian companies who are over age 50 (<https://valored.it/ricerche/talenti-senza-eta-2019/>). The sample here adopted consists of 9,575 employees (64.4% male) aged between 50 and 69 ($M = 55.33$; $SD = 3.31$). These employees belonged to 31 different companies. These companies had different sizes (from 13 to 73,727 employees), belonged to different business sectors (mainly chemical/pharmaceutical, food and beverage, energy/extractive industry, and media/entertainment) and were located in different Italian regions. On average, 309 employees filled in the questionnaire at each company (Min = 4; Max = 4815; $SD = 860.51$).

Measures

Social identity motives were measured using the scale proposed by Thomas et al. (2017). The scale is composed by six items because each item measures one of the six identity motives (self-esteem, distinctiveness, belonging, meaning, continuity and efficacy) that, according to the MICT (Vignoles, 2011), affect the identification with a group. Each item was evaluated on a 5-point scale (1 = *strongly disagree*; 5 = *strongly agree*). Sample items are: “I see [company name] as having a distinctive identity different from other companies” (distinctiveness motive); “I see [company name] having an identity that persists over time — from past to present to future” (continuity motive).

Data Strategy

MCFA, psychometric isomorphism, and level-specific reliability were analyzed using a 7-step procedure. More details can be found in Section “Results.”

Analyses were performed in Mplus 7.1 using the maximum likelihood estimation method. As missing data for each item of the scale ranged from 0.4% to 1.1% and they were completely at random — Little’s MCAR Test: $\chi^2(111) = 111.80$; $p = .461$ — the full information maximum likelihood was adopted as a strategy to manage the missing data. Mplus syntax for each step of the analysis can be found in Appendix.

Model fit was evaluated based on the following absolute fit indices that, according to Bonito and Keyton (2019), are the most recommended for MCFA: Chi-square test (χ^2), standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), and the comparative fit index (CFI).

Alternative models have to be compared (e.g., isomorphism testing) based on their absolute and relative fit indices (Bonito & Keyton, 2019). Specifically, in some studies (e.g., Tay et al., 2014), authors compared configural isomorphism model and metric isomorphism model using relative indices such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and sample size adjusted BIC (ssBIC). Models with the lowest information criteria, which indicate model parsimony, should be preferred. In other studies (e.g., Margola et al., 2019), authors compared configural and metric models using absolute indices. Specifically, a delta score (Δ) was calculated by subtracting CFI index of the metric model from CFI

index of the configural model. When ΔCFI is smaller than or equal to -0.01 , the two compared models can be considered equal, and the most parsimonious should be preferred (Cheung & Rensvold, 2002).

RESULTS

Testing the First Implicit Assumption: It is Possible to Measure the Construct at Both the Individual and Cluster Level

Step 1: Perform a Conventional Confirmatory Factor Analysis on the Sample Total Covariance Matrix

The first step consists of running a classic CFA, that is an analysis that ignores the nonindependence of data. In other words, in this step, participants are treated as independent individuals, not considering that they work for different companies. As stated by Dyer et al. (2005), results and fit indices of this model may be biased when there is appreciable nonindependence, especially when group sizes are large or when the factor structure at the within-group level of analysis is different from the factor structure at the between-group level (lack of isomorphism).

Expecting a unidimensional structure for the Social Identity Motives Scale (Thomas et al., 2017), we ran a CFA requiring that the six items loaded on the same latent factor. The obtained model had a partially satisfactory fit (see Table 1), with CFI and SRMR indicating perfect fit and RMSEA indicating lack of fit. This disagreement between indices may also depend on RMSEA, being the main index to be affected by the presence of too few items in the model (Chavez, Rodriguez, Vue, & Cabrera, 2019). Furthermore, Step 1 model had strong (from .771. to .892) and significant ($p < .001$) standardized factor loadings.

TABLE 1
Fit indices of measurement models

Model	χ^2	df	p	RMSEA [90% CI]	p	CFI	SRMR	AIC	BIC	ssBIC
Step 1	999.550	9	< .001	.107 [.102, .113]	< .001	.977	.018	122566.20	122695.20	122638.00
Step 3	880.977	9	< .001	.102 [.096, .108]	< .001	.977	.022	117734.42	117820.13	117781.99
Step 4a	197.152	9	< .001	.821 [.724, .923]	< .001	.697	.018	-274.760	-257.552	-294.938
Step 4b	197.298	10	< .001	.777 [.685, .874]	< .001	.698	.017	-276.614	-260.840	-295.111
Step 5	918.964	18	< .001	.072	-	.977	.022 (W) -.015 (B)	120142.245	120357.253	120261.92
Step 6	933.069	23	< .001	.064	-	.976	.022 (W) -.015 (B)	120146.351	120325.523	120246.08

Note. χ^2 = Chi-square; df = degree of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual; W = SRMR for the within-level; B = SRMR for the between-level; AIC = Akaike information criterion; BIC = Bayesian information criterion; ssBIC = sample-size adjusted BIC.

Step 2: Estimate Between-Group Level Variation

In this step, the proportion of systematic between-group variation for each observed variable in the model has to be calculated. In other words, there is the need to estimate how big the portion of the average participant's response that depends on his/her belonging to a specific company is. This estimation has to be done separately for each item of the scale.

Among the different analytic approaches useful to calculate this proportion of systematic between-group variation, intraclass correlation coefficient (ICC) is the most used (Lanz et al., 2018), but also design effect (DE) could be used to correct estimation for the average cluster size. Mplus automatically calculates ICC for each variable when the "type = twolevel basic" syntax is included in the "analysis" command, while DE should be calculated manually using the formula already presented in the present paper, that corrects ICC for c (average group size), which in our study is equal to 308.87. The cut-off for these two indices is respectively 0.05 and 2. As in our case (see Table 2), ICCs are higher than .09 and DEs are higher than 30, the need to perform a MCFA is evident.

TABLE 2
Indices of between-group variation for the six items of the Social Identity Motives Scale

Item	ICC	Design effect
1	0.120	37.945
2	0.128	40.407
3	0.127	40.100
4	0.096	30.556
5	0.136	42.870
6	0.129	40.715

Note. ICC = intraclass correlation coefficient.

Testing the Second Implicit Assumption: The Construct at the Cluster Level has the Same Dimensions as the Construct at the Individual Level (Configural Isomorphism)

Step 3: Perform a Factor Analysis on the Sample Pooled-Within Covariance Matrix

If in the second step participants' responses depended on both the participants themselves (within-level) and the group to which they belong (between-level), researchers have to analyze the factorial model separately for within-level (Step 3) and between-level (Step 4) covariance matrices.

Even if Steps 1, 3, and 4 could seem similar as they test the same factorial model on just one level, an important difference encompasses these three steps and it consists of the covariance matrix they use to estimate the model. As stated before, when a classic CFA model is run (like Step 1) the total sample covariance matrix (S_T) is used and, consequently, the within- and between-level variances are not distinguished. The S_T can be decomposed in two other matrices: S_{PW} and S_B , containing respectively the within-level variation and the across-group variation. In Step 3 the same factorial structure tested in Step 1 is estimated based on the S_{PW} , that is accounting only for the within-level variation. Vice versa, in Step 4, the same factorial structure is estimated based on the S_B , that is accounting only for the between-group variation. Step 3 and

Step 4 are essential as they give information that are level-specific and are precursors of Step 5 where both levels are tested separately but simultaneously.

Thus, to perform Steps 3 and 4, researchers have to extract within- and between-level covariance matrices from the sample total covariance matrix (see syntax in Appendix). Using the within-level covariance matrix, researchers perform the factor analysis on the sample pooled-within covariance matrix (Step 3) and compare it with the factor analysis obtained in the first step. The model estimated using the within-level covariance matrix may show a worse fit if the predominance of the construct-relevant variance is at the between-group level. On the other hand, the model estimated using within-level covariance may show an improved fit if the construct relevant variance is primarily at the within-group level, or the factor structures differ substantially at the two levels.

In our case (see Table 1), the Step 3 model is only slightly improved with respect to the Step 1 model. Standardized factor loadings of this factor model were all strong (from .743. to .875) and significant ($p < .001$).

Step 4: Perform a Factor Analysis on the Sample Between-Group Covariance Matrix

The Step 3 analysis has to be repeated but using the between-group covariance matrix (Step 4). This model had very low fit indices (see Table 1, Step 4a) as well as an estimation problem: the residual variance of Item 5 was negative (0 for the unstandardized solution and $-.001$ for the standardized solution). As suggested by Dyer et al. (2005), we fixed Item 5 error variance to zero in order to avoid a negative variance estimate. The new model (see Table 1, Step 4b) had no estimation problems, but its fit indices remained substantially poorer than the Step 3 model (the decreased fit is considered almost usual; Dyer et al., 2005).

The Chi-square improved due to the substantially smaller sample size used in Step 4 (i.e., the number of companies vs. the number of employees). Standardized factor loadings of this factor model were all very strong (from .940. to 1.00) and significant ($p < .001$).

Step 5: Perform Multilevel Confirmatory Factor Analysis (Configural Isomorphism)

In Step 5, researchers have to perform the MCFA considering both the within- and the between-level variation of the responses. In other words, in this step the expected factorial structure is simultaneously tested at both within- and between-level, basing the estimation respectively on within- and between-level variation. If researcher assumes that the construct at the cluster level has the same dimensions as the construct at the individual level, the same number of factors (one factor in our example) and the same equivalence in the zero-and-nonzero loading patterns will be required at both levels. If fit indices of this model are good, researchers can conclude that there is configural invariance across the two levels, that is, the construct has the same dimensions at both individual and cluster level.

As shown in Table 1, the Step 5 model is the one that fits our data best. At both levels, all factor loadings were significant at $p < .001$ and were strong at both the within (standardized factor loadings ranging from .743 to .875) and between (standardized factor loadings ranging from .937 to .998) levels. Note that in this model Item 5's residual variance was not fixed to 0 and all the standardized factor loadings were freely estimated.

Good fit indices of this model support the configural isomorphism of the Social Identity Motives Scale. In other words, the social identity construct has the same factorial structure (unidimensional structure) at both the individual and company levels, as hypothesized by Thomas et al. (2017).

Testing the Third Implicit Assumption: The Construct at the Cluster Level has the Same Meaning of the Construct at the Individual Level (Metric Isomorphism)

Step 6: Metric Isomorphism

In the sixth step, metric isomorphism was tested to verify if the higher-level construct has a similar meaning as its lower-level counterpart (Tay et al., 2014). If Step 5 verifies that the number of dimensions is invariant across levels, Step 6 verifies that the association between each item and each dimension (i.e., factor loading) is invariant across level. If item-dimension association is the same, it means that the content of each items describes the dimension equally well on the two levels and that, consequently, the dimension has the same meaning on the two levels.

To test that assumption, it is necessary to set factor loadings the same across levels (Figure 1) and compare the fit indices of this constrained model (Step 6) with the unconstrained configural model (Step 5). Specifically, when comparing *relative* fit indices (AIC, BIC, and adjBIC), two indices out of three suggested that the constrained model (Step 6) is better than the unconstrained one (Step 5). The same conclusion is achieved comparing the models through *absolute* fit indices. Specifically, the ΔCFI , being equal to $-.001$ (obtained subtracting $.976$ to $.977$), is below the $-.01$ cut-off suggested by Cheung and Rensvold (2002).

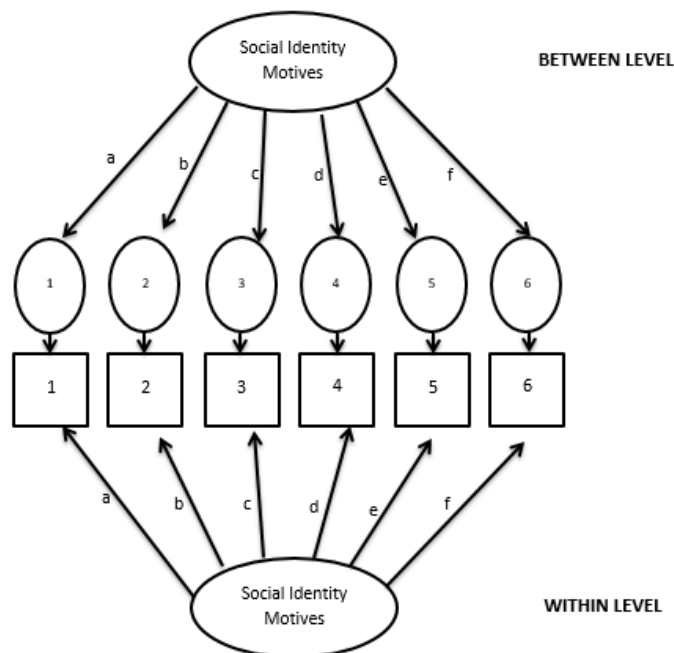


FIGURE 1
Metric isomorphism model of the Social Identity Motives Scale.

Testing the Fourth Implicit Assumption: Reliability is the Same at the Cluster and Individual Levels
(Level-Specific Reliability)

Step 7: Level-Specific Reliability

Once the most plausible factorial structure is identified, the internal consistency of the factor(s) at each level has to be tested in order to estimate the instrument's reliability. It is important to test reliability after testing metric isomorphism so that researchers can include constraints put in the factorial structure found in the previous step. In our case, we tested within- and between-level reliability keeping the factor loadings' constraints across levels. Specifically, we tested omega (ω ; Geldhof et al., 2014). Results suggest that the Social Identity Motives Scale is strongly reliable at both within ($\omega = .919$) and between ($\omega = .998$) levels.

DISCUSSION

This paper's aim was to illustrate a seven-step approach useful for explicitly testing four implicit assumptions that researchers often make when they aggregate Level 1 scores (e.g., employees' scores) into a Level 2 score (e.g., company's score). Specifically, this paper presents a procedure for performing MCFA in order to explicitly assume that there is variance explained by the cluster level, configural, and psychometric isomorphism, and acceptable reliability at both cluster and individual levels.

That multistep procedure was applied to test the measurement model of the Social Identity Motives Scale that measures an individual level (i.e., perception that individuals have of the group they belong) and a group level construct (i.e., average perception of group members). In the original paper, Thomas et al. (2017) implicitly assumed that the scale was isomorphic across levels, as they operationalized higher-level construct using an aggregated of individual-level scores within groups. However, that aggregation is based on an implicit assumption: within- and between-level constructs have the same ontology (Jak & Jorgensen, 2017). Many researchers have that implicit belief, but they are not aware that the cross-level ontology of multilevel construct is not something obvious, but needs to be proved.

In the present paper, we adapted the scale to the organizational context, measuring employees' social identity motives at the individual and cluster levels, and testing the four implicit assumptions through the seven-step procedure. Steps 1 and 2 allowed for testing that the Social Identity Motives Scale items' variances are explained by both between and within levels. Steps 3, 4, and 5 allowed testing configural isomorphism. The sixth step allowed us to prove that the meaning of the construct was the same at within and between levels (psychometric isomorphism testing); the seventh step (level-specific reliability) showed that the latent factor scores are highly reliable both at the within and between levels. In order to go into the meaning of those findings, we now discuss what to do when implicit assumptions are verified, and when they are not.

When Implicit Assumptions are Verified

When the first assumption ("It is possible to measure the construct at both the individual and cluster level") is verified (ICC higher than .05 and DE higher than 2), researchers need to go on and test for psychometric isomorphism and level-specific reliabilities. Indeed, they are aware that individual scores are influenced by their belonging to different clusters (i.e., companies), so that between factor(s) in the MCFA have to be considered.

When the second (“The construct at the cluster level has the same dimensions as the construct at the individual level”) and third (“The construct at the cluster level has the same meaning as the construct at the individual level”) assumptions are verified, psychometric isomorphism (i.e., both configural and metric isomorphism) is found. Jak and Jorgensen (2017) stated that psychometric isomorphism is a prerequisite to operationalizing higher-level constructs as aggregation of individual-level scores within groups. For example, if we measure experience across different employees belonging to different companies, we can operationalize the construct “experience” at the company level by aggregating the scores of the members who belong to the same company only if psychometric isomorphism has been found.

Moreover, psychometric isomorphism is also a prerequisite for testing score or nomological similarity (Tay et al., 2014). Score similarity consists of verifying how similar scores belonging to the same cluster are (e.g., how similar scores of employees belonging to the same company are), and assumes psychometric isomorphism across levels. Nomological similarity across-level (or homology) verifies to what extent the relations across different constructs are equivalent across levels (Chen, Bliese, & Mathieu, 2005). However, homology can be established only if psychometric isomorphism has been demonstrated. For example, researchers can compare the relationship between performance and social identity motives at the within-level with the same relationship at the between-level only if the two constructs have showed psychometric isomorphism (i.e., if the two constructs have the same meaning/nature at both levels). A lack of psychometric isomorphism leads to invalid tests of homology. In other words, in order to compare the relationship that two variables have at the between-level with the same relationship at the within-level, researchers need to verify that the two variables have the same meaning across levels (psychometric isomorphism). Indeed, from a construct validation standpoint, it is well-known that measurement equivalence needs to be established before predictive equivalence (Vandenberg & Lance, 2000). In sum, finding both configural and metric isomorphism allows researchers to aggregate individual-level scores measured within groups into a between-level score, calculate reliability at both the within and between levels, and test homology across levels.

Finally, when the fourth assumption (“The reliability is the same at the cluster and individual levels”) is also verified, researchers can affirm that residual variances are not determining a lower reliability related to the between- or within-level score. In this way, the aggregate variable can be considered as reliable as the within-level variable.

When Implicit Assumptions are not Verified

On the other hands, it is important considering what researchers can do when they do not confirm the four assumptions. If items’ between variance (first assumption) is not verified (i.e., ICCs values lower than 0.05), the nested structure of the data is not affecting the individual responses. For example, we estimated ICC for the items of Flourishing Scale (Diener et al., 2010), an instrument that assesses the individual psychological well-being and that was administered to the employees of our research project. ICCs values for the eight items of the scale ranged from 0.01 to 0.02. As those values were lower than the cut-off, we can conclude that the individual psychological well-being employees reported was not affected by the company they belonged. In this case, classical CFA is allowed and no estimation bias due to the nested design will be present. Moreover, it also means that clusters (i.e., companies) are not different in the average perception of their members (i.e., employees), so between-level variables should not be computed.

If between-level variance is verified, but not configural isomorphism, researchers should test at least a weak configural isomorphism (Tay et al, 2014). Weak configural isomorphism consists of having the same

number of factors across levels without equivalence in the zero-and-nonzero loading patterns. If weak configural isomorphism is found, researchers should question the similarity of the latent factors' definition. However, it is difficult to obtain weak configural isomorphism because, in most of the cases in which strong configural isomorphism is not found, the between-level shows a lower number of latent factors than the within-level (e.g., Margola et al., 2019). It seems that in those kinds of models, item variance at the between level can be explained by fewer latent factors than the within-level variance. Researchers should reflect on the implication that the absence of configural isomorphism has for the theoretical model of the investigated construct. For instance, Margola and colleagues (2019) showed that the Family Relationships Index (FRI) — an instrument measuring the quality of family relationships — has three dimensions at the individual level (conflict, cohesion, communication) and just one dimension at the family level (relationship quality). Theoretically speaking, we could speculate that at family level the quality of relationships is conceptualized more as a homogeneous construct, resulting in one dimension, whereas, at the individual level, family members differentiate among different facets of that quality.

In other cases, configural isomorphism is verified, but not the equivalence of factor loadings across levels (lack of metric isomorphism; Tay et al., 2014). In those cases, researchers should first try to test a partial metric isomorphism model. This means progressively releasing factor loadings' constraints in order to determine which parameters did not meet isomorphism across levels. For instance, Ruelens, Meuleman, and Nicaise (2018) found that two of five indicators (Trust in National Parliament, Politicians, Political Parties, Legal System and Police) measuring political trust were not equal across levels. In particular, findings showed that political trust was conceptualized more as based on trust in legal system and police at country level than at the individual level. When the partial metric isomorphism model is not significantly different from the configural isomorphism model, researchers should interpret the latent factor differently across levels. Even if the latent factor is measured using the same items, the absolute weight that some items have in defining the construct changes across levels. Researchers should consider items that are weighted more (i.e., have higher factor loadings) as closer to the real meaning of the measured construct at that level.

If partial metric isomorphism is not found, researchers should verify if they can find at least a weak form of metric isomorphism (Tay et al., 2014). Weak metric isomorphism is found when the relative ordering of loadings (not their absolute values) is equivalent between levels. In this case, the meaning of the construct remains similar across levels (items that carry more weight in defining the construct are the same), but the item's residual variance is different across levels. This means that the level that has the smaller factor loadings in absolute value terms is the one that measure the construct with higher measurement error.

Finally, if all the three previous assumptions are found, but reliability is not good at the between-level, researcher should consider aggregating individual-level scores excluding the residual error, which means saving a factor score at the between-level. Indeed, reliability is computed by considering residual errors at both within and between levels, and thus low reliability when metric isomorphism is found is probably due to issues regarding residual errors (Geldhof et al., 2014).

A Last Implicit Assumption About the Content of the Construct at the Cluster Level

Recent literature underlines a last implicit assumption about the conceptual interpretation of cluster-level constructs. According to Stapleton and Johnson (2019), the between-level factor isomorphic to the within-level factor should be interpreted as the average score of the construct reported by members belonging to the same group, as we suggested in the present paper. In that case, the construct is a property of the cluster's

members, not a property of the cluster. Thus, for example, if the construct is the employee's experience within a specific company, the between-level factor will be the experience of the "average employee" of that company. We can use the same reasoning when the construct is referred to as a property of the cluster, such as a property of the company. Thus, for example, employees are asked to evaluate the performance or the climate of their company. The implicit assumption is that the between-level construct is measuring a property of the company, such as the company's performance. However, we should test that implicit assumption. Indeed, it could also be that we are only measuring the "average employee" 's perception of the company's performance. Thus, it is not sufficient to conduct *isomorphic* MCFA as we did in the present paper, but it is important to test a specific model, called the *shared/isomorphic* MCFA model (Bonito & Keyton, 2019; Stapleton & Johnson, 2019). If the aim is to test the company's performance or climate, researchers need to explicitly assume that, at the between-level, two different constructs are measured: the average of the employees' perception of the company's climate or performance (configural factor) and the shared factor that is the variance not explained by the average perception and is interpreted as the company's climate or performance. The shared factor has to be added to the between-level in addition to the latent factor isomorphic to the within-level latent factor. Note that the shared factor should be always hypothesized when fit indices of the *isomorphic* MCFA are not good (Stapleton & Johnson, 2019).³

Limitations

The current paper has been proposed as a tutorial for researchers aiming to test measurement models of nested data in order to verify their implicit measurement assumptions. In order to facilitate the comprehension of the proposed seven steps, we used a simple illustration based on a brief and unidimensional scale. In research practice, researchers may have to use longer scales where items are organized across different factors. This could create some difficulties in finding configural and metric isomorphism. Furthermore, our data were nested at one level (company), while other researchers may have data nested across more levels (e.g., employees nested both in teams and companies; Liu et al., 2011). This requires testing psychometric isomorphism across more than two levels. Finally, we presented how to test the measurement model of a scale adopting classic test theory as a psychometric framework, but other researchers could adopt item response theory, requiring differences in the approach testing for configural and metric isomorphism (Tay et al., 2014).

Even if our paper has those limitations, it has the advantage of working well as introduction to topics such as MCFA and isomorphism that many authors are not familiar with. All the disciplines testing multilevel models using aggregate scores (Alm et al., 2019) could benefit from the steps we proposed here (e.g., Stapleton and Johnson, 2019 proposed solutions for the educational literature).

In sum, our suggestion is that researchers collecting nested data should always consider testing the measurement model of their construct using a *isomorphic MCFA*. Furthermore, researchers should verify the assumptions regarding measurement and meaning of the construct and, according to the results obtained on psychometric isomorphism, they may reach different conclusions that have different practical implications; the same is true regarding aggregation.

NOTES

1. For the sake of simplicity, from here on we speak about multilevel structure based on two levels (e.g., employee and company), but multilevel analysis can be also applied to nested structure composed by more than two levels (e.g., employees, teams, companies; see e.g., Liu, Chen, & Yao, 2011).

2. Tay et al. (2014) used the expression cross-level isomorphism to indicate the condition in which “higher-level constructs have similar meanings and properties as their lower-level counterparts” (p. 78). According to Tay et al., isomorphism in multilevel constructs, in its ideal form, involves three aspects: (1) psychometric isomorphism across levels; (2) nomological similarity across levels; and (3) score similarity within collective units. In the current paper we are going to present only psychometric isomorphism, being that our focus is on the measurement model of multilevel constructs.
3. For further details, see Stapleton and Johnson (2019)’s paper that also offer Mplus syntax to run this shared/isomorphic MCFA model.

FUNDING

The survey utilized in this study was financed by Valore D (<https://valored.it/en/>). The publication of the paper is funded by a grant from Catholic University.

REFERENCES

- Alm, S., Låftman, S. B., Sandahl, J., & Modin, B. (2019). School effectiveness and students’ future orientation: A multilevel analysis of upper secondary schools in Stockholm, Sweden. *Journal of Adolescence*, 70, 62-73.
doi:10.1016/j.adolescence.2018.11.007
- Bentler P. M. (2009). Alpha, dimension-free and model-based internal consistency reliability. *Psychometrika*, 74(1), 137-143.
doi:10.1007/s11336-008-9100-1
- Bonito, J. A., & Keyton, J. (2019). Multilevel measurement models for group collective constructs. *Group Dynamics: Theory, Research, and Practice*, 23(1), 1-21.
doi:10.1037/gdn0000096
- Chavez, C., Rodriguez, M. C., Vue, K., & Cabrera, J. (2019, April). *A validity argument sensitivity analysis of social and emotional learning measures with few items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, NY. Retrieved from <https://conservancy.umn.edu/bitstream/handle/11299/203637/2019-few-items-NCME.pdf?sequence=1>
- Chen, G., Bliese, P. D., & Mathieu, J. E. (2005). Conceptual framework and statistical procedures for delineating and testing multilevel theories of homology. *Organizational Research Methods*, 8, 375-409.
doi: 10.1177/1094428105280056
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
doi:10.1207/S15328007SEM0902_5
- Claxton, S. E., DeLuca, H. K., & Van Dulmen, M. H. (2015). Testing psychometric properties in dyadic data using confirmatory factor analysis: Current practices and recommendations. *Testing, Psychometrics, Methodology in Applied Psychology*, 22(2), 181-198.
doi:10.4473/TPM22.2.2
- Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children’s mental health agencies. *Journal of Emotional and Behavioral Disorders*, 19(1), 27-40.
doi:10.1177/1063426610365879
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97(2), 143-156.
- Dixon, M. A., & Cunningham, G. B. (2006). Data aggregation in multilevel analysis: A review of conceptual and statistical issues. *Measurement in Physical Education and Exercise Science*, 10(2), 85-107.
- Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, 16(1), 149-167.
doi:10.1016/j.leaqua.2004.09.009
- Elosua, P., & Hermosilla, D. (2013). Does body dissatisfaction have the same meaning for males and females? A measurement invariance study. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 63(5), 315-321.
doi:10.1016/j.erap.2013.06.002
- Forer, B., & Zumbo, B. D. (2011). Validation of multilevel constructs: Validation methods and empirical findings for the EDI. *Social Indicators Research*, 103(2), 231-265.
doi:10.1007/s11205-011-9844-3

- Garson, G. D. (2013). Introductory guide to HLM with HLM 7 software. In G. D. Garson (Ed.), *Hierarchical linear modeling: Guide and applications* (pp. 55-96). Thousand Oaks, CA: Sage.
- Geiser, C. (2013). *Data Analysis with Mplus*. New York, NY: The Guilford Press.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72-91.
doi:10.1037/a0032138
- Jak, S., & Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Frontiers in Psychology*, 8, 1640.
doi:10.3389/fpsyg.2017.01640
- Lanz, M., Sorgente, A., & Tagliabue, S. (2018). Inter-rater agreement indices for multiple informant methodology. *Marriage & Family Review*, 54(2), 148-182.
doi:10.1080/01494929.2017.1340919
- Li, F., Duncan, T. E., Harmer, P., Acock, A., & Stoolmiller, M. (1998). Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(3), 294-306.
doi:10.1080/10705519809540106
- Liu, D., Chen, X. P., & Yao, X. (2011). From autonomy to creativity: A multilevel investigation of the mediating role of harmonious passion. *Journal of Applied Psychology*, 96(2), 294-309.
doi:10.1037/a0021294
- Margola, D., Fenaroli, V., Sorgente, A., Lanz, M., & Costa, G. (2019). The Family Relationships Index (FRI): Multilevel confirmatory factor analysis in an Italian community sample. *European Journal of Psychological Assessment*, 35, 335-345.
doi:10.1027/1015-5759/a000427
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1), 85-112.
- Raykov T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173-184.
doi: 10.1177/01466216970212006
- Ruelens, A., Meuleman, B., & Nicaise, I. (2018). Examining measurement isomorphism of multilevel constructs: The case of political trust. *Social Indicators Research*, 140(3), 907-927.
- Satorra, A., & Muthen, B. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25(1), 267-316.
doi:10.2307/271070
- Sorgente, A., Tagliabue, S., & Lanz, M. (2019). Brief Inventory of Thriving: Testing Italian version's reliability using SEM. *Psicologia della Salute*, 21, 97-117.
doi:10.3280/PDS2019-001006
- Stapleton, L. M., & Johnson, T. L. (2019). Models to examine the validity of cluster-level factor structure using individual-level data. *Advances in Methods and Practices in Psychological Science*, 2(3), 312-329.
- Suzuki, E., Yamamoto, E., Takao, S., Kawachi, I., & Subramanian, S. V. (2012). Clarifying the use of aggregated exposures in multilevel models: self-included vs. self-excluded measures. *PloS One*, 7(12), e51717.
doi:10.1371/journal.pone.0051717
- Tagliabue, S., & Lanz, M. (2014). Exploring social and personal relationships: The issue of measurement invariance of non-independent observations. *European Journal of Social Psychology*, 44(7), 683-690.
doi:10.1002/ejsp.2047
- Tay, L., Woo, S. E., & Vermunt, J. K. (2014). A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organizational Research Methods*, 17(1), 77-106.
doi:10.1177/1094428113517008
- Thomas, W. E., Brown, R., Easterbrook, M. J., Vignoles, V. L., Manzi, C., D'Angelo, C., & Holt, J. J. (2017). Social identification in sports teams: The role of personal, social, and collective identity motives. *Personality and Social Psychology Bulletin*, 43(4), 508-523.
doi:10.1177/0146167216689051
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.
doi:10.1177/109442810031002
- Vignoles, V. L. (2011). Identity motives. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 403-432). Springer Science & Business Media.
doi:10.1007/978-1-4419-7988-9_18
- Vignoles, V. L., & Manzi, C. (2014). Identity motives. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3061-3064). Dordrecht, Netherlands: Springer.
doi:10.1007/978-94-007-0753-5_3312
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123-133. doi:10.1007/s11336-003-0974-7

APPENDIX

Syntaxes to Run Each Step

Step 1: Perform a Conventional Confirmatory Factor Analysis on the Sample Total Covariance Matrix

DATA:

FILE IS identity.dat;

VARIABLE:

NAMES ARE

ID

company

item1

item2

item3

item4

item5

item6;

USEVARIABLES ARE

item1

item2

item3

item4

item5

item6;

MISSING ARE all (999);

MODEL:

SIM BY

item1

item2

item3

item4

item5

item6;

OUTPUT:

STAND SAMP MOD;

Step 2: Estimate Between-Group Level Variation

DATA:

FILE IS identity.dat;

VARIABLE:

NAMES ARE

ID

company

item1

item2

item3

item4

item5

item6;

USEVARIABLES ARE

item1

item2

item3

item4

item5

item6;

MISSING ARE all (999);

CLUSTER IS company;

ANALYSIS:

TYPE= TWOLEVEL BASIC;

OUTPUT:

SAMP;

SAVEDATA:

SIGBETWEEN IS BetCov.dat;

SAMPLE IS WinCov.dat;

Step 3: Perform a Factor Analysis on the Sample Pooled-Within Covariance Matrix

DATA:

FILE IS WinCov.dat;

FORMAT IS free;

TYPE IS COVA;

NOBSERVATION IS 9344; *!report here number of participants (9575) less number of groups (31)*

VARIABLE:

NAMES ARE *!include here only variables that were listed in Step 2 "USEVARIABLES ARE"*

item1
item2
item3
item4
item5
item6;

MODEL:

SIM BY

item1
item2
item3
item4
item5
item6;

OUTPUT:

STAND SAMP MOD;

Step 4: Perform a Factor Analysis on the Sample Between-Group Covariance Matrix

DATA:

FILE IS BetCov.dat;
FORMAT IS free;
TYPE IS COVA;
NOBSERVATION IS 31; *!report here the number of groups (31 companies)*

VARIABLE:

NAMES ARE *!include here only variables that were listed in Step 2 "USEVARIABLES ARE"*

item1
item2
item3
item4
item5
item6;

MODEL:

SIM BY

item1
item2
item3
item4
item5
item6;

item5@0; *!this line was added in step 4b*

OUTPUT:

STAND SAMP MOD;

Step 5: Perform Multilevel Confirmatory Factor Analysis (Configural Isomorphism)

DATA:

FILE IS identity.dat;

VARIABLE:

NAMES ARE
ID

company
item1
item2
item3
item4
item5
item6;

USEVARIABLES ARE

item1
item2
item3
item4
item5
item6;

MISSING IS ALL (999);

CLUSTER IS company;

ANALYSIS:

TYPE IS TWOLEVEL;
ITERATIONS IS 5000;
ESTIMATOR IS ML;

MODEL:

%BETWEEN%

B_SIM BY

item1
item2
item3
item4
item5
item6;

%WITHIN%

W_SIM BY

item1
item2

item3
item4
item5
item6;

OUTPUT: STAND SAMP MOD;

Step 6: Metric Isomorphism

DATA:

FILE IS identity.dat;

VARIABLE:

NAMES ARE

ID

company

item1

item2

item3

item4

item5

item6;

USEVARIABLES ARE

item1

item2

item3

item4

item5

item6;

MISSING IS ALL (999);

CLUSTER IS company;

ANALYSIS:

TYPE IS TWOLEVEL;

ITERATIONS IS 5000;

ESTIMATOR IS ML;

MODEL:

%BETWEEN%

B_SIM BY

item1 (a)

item2 (b)

item3 (c)

item4 (d)

item5 (e)

item6 (f);

%WITHIN%

W_SIM BY

item1 (a)
item2 (b)
item3 (c)
item4 (d)
item5 (e)
item6 (f);

OUTPUT: STAND SAMP MOD;

Step 7: Level-Specific Reliability

DATA:

FILE IS identity.dat;

VARIABLE:

NAMES ARE

ID
company
item1
item2
item3
item4
item5
item6;

USEVARIABLES ARE

item1
item2
item3
item4
item5
item6;

MISSING IS ALL (999);

CLUSTER IS company;

ANALYSIS:

TYPE IS TWOLEVEL;
ITERATIONS IS 5000;
ESTIMATOR IS ML;

MODEL:

%BETWEEN%

B_SIM BY

item1* (a)
item2 (b)
item3 (c)
item4 (d)
item5 (e)

item6 (f);

B_SIM@1;

item1 (g);
item2 (h);
item3 (i);
item4 (l);
item5 (m);
item6 (n);

%WITHIN%
W_SIM BY
item1* (a)
item2 (b)
item3 (c)
item4 (d)
item5 (e)
item6 (f);

W_SIM@1;

item1 (o);
item2 (p);
item3 (q);
item4 (r);
item5 (s);
item6 (t);

OUTPUT: SAMP;

MODEL CONSTRAINT:

NEW(NUM DEN_B OMEGA_B DEN_W OMEGA_W);

NUM = (a+b+c+d+e+f)**2; *!numerator . It is the same for the between level and the within
! level only if the metric invariance is present*

DEN_B = ((a+b+c+d+e+f)**2)+(g+h+i+l+m+n); *!denominator – between level*
OMEGA_B = NUM/DEN_B;

DEN_W = ((a+b+c+d+e+f)**2)+(o+p+q+r+s+t); *!denominator – within level*
OMEGA_W = NUM/DEN_W;