Reflective vs. Formative Measurement Models in Operations and Supply Chain Research

*Pietro De Giovanni**

Received December 15, 2024 – Accepted February 20, 2025

Abstract

This research seeks to highlight a common mistake that researchers in the area of Operations and Supply Chain Management (O&SCM) make when selecting the measurement models in Structural Equation Modelling. In fat, the unproper selection of a measurement model in Structural Equation Modeling (SEM) research can lead to issues of model misspecification and non-valid findings. Therefore, this is the first study in O&SCM that highlights the differences between reflective and formative measurement models in SEM and invites researchers in this field to reflect and pay attention to the measurement model selection before diving into a statistical analysis.

Keywords: Structural Equation Modeling, Formative Models, Reflective Models, Misspecification.

Modelli di misurazione Riflessivi vs. Formativi in Operations e Supply Chain

Sommario

Questa ricerca si propone di evidenziare un errore comune commesso dai ricercatori nel campo dell'Operations e Supply Chain Management (O&SCM) nella

* SDA Professor. SDA Bocconi School of Management, Strategy & Operations Knowledge Area. pietro.degiovanni@sdabocconi.it

Corporate Governance and Research & Development Studies, n. 1/2025
(ISSN 2704-8462-ISSNe 2723-9098, Doi: 10.3280/cgrds1-2025oa19037)

scelta dei modelli di misurazione all'interno della Structural Equation Modeling (SEM). Infatti, una scelta inappropriata del modello di misurazione nella ricerca basata sulla SEM può condurre a problemi di misspecification del modello e a risultati non validi. Pertanto, questo è il primo studio nel campo dell'O&SCM che mette in luce le differenze tra i modelli di misurazione riflessivi e formativi nella SEM, invitando i ricercatori in questo filone a riflettere e a prestare la dovuta attenzione alla selezione del modello di misurazione prima di procedere con l'analisi statistica.

Parole chiave: Structural Equation Modeling, Modelli Formativi, Modelli Riflessivi, Specificazione del modello.

1. Introduction

Recent research in various fields has highlighted the scarce attention in choosing the proper measurement model in Structural Equation Modelling (SEM) (e.g., Jarvis *et al*., 2003; Petter *et al.*, 2008). The correct correspondence between constructs and measures is a needed prerequisite to develop theories through a correct research hypotheses testing and thus avoid misspecification concerns. The classical test theory assumes causality between a construct and its measures, thus a change in the construct *causes* a change in the related measures (Bollen, 1989). Although this assumption holds in many instances, it turns out to be inappropriate in all cases when a construct can be seen an index made by some observable variables rather than being considered as their cause. Technically, this difference consists of the right selection of the measurement models, which can be either reflective or formative. Formally, the model misspecification leads to several problems such as: different assumptions, different interpretation of relationships and, most importantly, different statistical results.

Choosing the measurement can lead to serious concerns. As highlighted by Jarvis *et al*. (2003), 29% the articles in marketing has wrongly selected the measurement models for their empirical studies. Similarly, Petter *et al*. (2008) highlighted the common mistake that empirical research shares in the field information systems: a three-year analysis of the literature has revealed the scarce attention paid by researchers in the analysis and selection of measurement models. Podsakoff *et al*. (2006) show inappropriate modeling for 62% of constructs published in three major strategic management journals, while Podsakoff *et al*. (2003) report a misspecification rate of 47% for leadership research. In this research, I wish to put a lent on the same concern for Operations and Supply Chain Management (O&SCM) research.

The use of reflective measurement seems to be more a methodological tradition than being supported by robust theoretical motivations. For instance, all constructs related to the various forms of O&SCM (e.g., Operational practice, SC Integration, SC Management) have used reflective measures while Yeung (2008) used formative indicators to model O&SCM construct. A few questions need to be answered:

"Which approach (reflective or formative) should researchers then adopt? Which of these is the most appropriate? How can technical problems be overcome?".

In this paper, I wish to put light into this type of decisions, in which the usage of reflective model is either a tradition or a way for escaping some technical restrictions due to a covariance-based approach (Lisrel). I supply a list of criteria and suggestions that researchers in O&SCM domain can follow when they set up the conceptual and the measurement models. I also provide an example in Green SC research to demonstrate that a correct model specification leads to correct findings.

The paper is organized as follows. In section 2, I develop a literature review on the use of measurement models in the field of O&SCM. Afterwards, in section 3, I introduce the differences between formative and reflective models and highlight the technical concerns related to their application. In section 4, I provide an example of misspecification problems as well as an example, while section 5 concludes reporting both managerial insights and research advices.

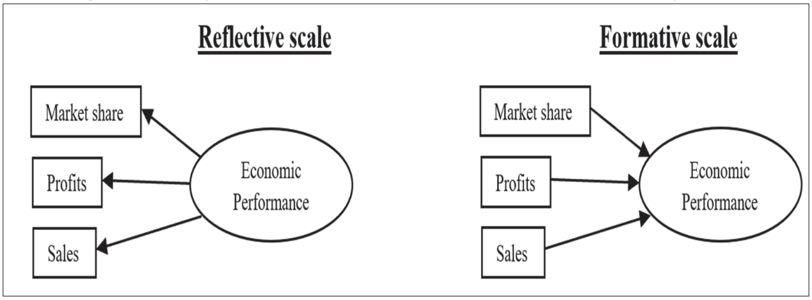
2. Literature Review and Reflections on the Measurement Model

Empirical research in O&SCM using SEM has mainly focused on the use of reflective scales to model relationships between a construct and its underlying measures. Notwithstanding, researchers should start modeling a problem looking into the relationships between constructs and variables. Although the answer for this question seems to be extremely easy for some constructs, it is not always the case for some others. Conceptually, one should think at the problem: *who comes first? Chicken or eggs?* Therefore, the selection of the right model passes through a similar question, that is: *who comes first? The construct or the manifest variables?*

When the latent variable comes for first, it exists independently of its measures. In such a case, a reflective model should be selected. Instead, the manifest variables come first when the construct does not exist without these items. In such a case, the model is formative because each measure contributes to *form* the construct, which is now called composite variable.

To better understand the selection process, I take the example of the construct Economic Performance, which is commonly used in O&SCM research as an endogenous construct. Figure 1 displays the construct measured through a reflective and a formative model in which the indicators are market share, profits, and sales. These measures of Economic Performance are really common in O&SCM research. Yet, the question that researchers should ask is: *Which model should one select?*

Figure 1 – Reflective vs. Formative measurement models for EP



Author's development

The literature in O&SCM has used both approaches described in Figure 1. For example, De Giovanni (2020), Roh *et al*. (2022), Alsheyadi *et al*. (2024) and many others used reflective models to measure Economic Performance. Differently, Chand *et al*. (2022), Malesios *et al*. (2020), De Giovanni and Vinzi (2012), De Giovanni (2012), Xu *et al*. (2019), Wallenburg and Weber (2005) and Garver (2019) used the formative models to measure Economic Performance. To properly choose the model, one should keep in mind that the directions of arrows alone are not at all sufficient to clearly identify the most appropriate model. Rather, researchers should better think about the meaning of a construct, what it does represent, what its role is inside the conceptual model, and what type of information one would obtain. Any time research seeks to investigate Economic Performance as a global indicator of economic value that a firm gains in a given instant of time, a formative model should be selected. In this case, in fact, profits, market share, and sales provide a contribution to form the firms' economic performance. With the elimination of one of these variables, for instance, profits, I do eliminate an important dimension of Economic Performance and the meaning of the construct may completely change.

However, most of the research in O&SCM studied performance as a reflective latent variable. In this case, a reflective measurement model allows the investigation of a soft and intangible asset, which a firm possesses independently of any other factor, for instance, the *Firms' Capability to create Economic Value*. In such case, the meaning of the construct is absolutely different from the concept emerged when using a formative measurement model.

In fact, a reflective measurement model studies a capability, an attitude, an ability to produce satisfactory economic outcomes. The firm obtains some profits because it possesses the capability to perform from an economic perspective. Therefore, causality goes from the construct to the measure and a reflective scale is the most appropriate. The construct comes before the measures as it exists even without its measures. In contrast, a formative measurement model allows the identification of a global indicator of Economic Performance, in which the construct is a composite index. The firm`s economic results are determined by its indicators and causality goes from the measures to the construct. The measures come before the construct as the latter does not exist without its measures.

Table 1 – Types of errors in scale selection

		Measurement model needed	
		Reflective	***Formative***
Measurement model adopted	***Reflective***	Correct model specification	Error of Type I
	Formative	Error of Type II	Correct model specification

Author's development

These distinctions can make a great difference in O&SC studies. While a conceptual and theoretical discussion should precede the model selection, researchers need to identify and clearly state the purposes of each construct inside the research. Not only the interpretation of a construct is different, but also the research hypotheses characterization as well as the results obtained in the confirmatory and structural analyses. Many problems of misspecification may occur at this stage. Diamantopoulos and Siguaw's (2006) referred to Error of Type I and II the adoption of erroneous measurement models. Specifically, the adoption of reflective indicators where formative indicators would be appropriate leads to Error of Type I; the adoption of formative indicators where reflective indicators would be appropriate leads to Error of Type II.

The identification of these misspecification cases affects on model estimates and fit statistics, thus influencing the conclusions about the theoretical relationships among the constructs drawn from the research (Jarvis *et al.*, 2003). Only a few papers made a reflection on the measurement model selection in O&SCM. For example, Chand *et al*. (2022) measure the impact of complexity on performance by only using formative measurement models for the constructs: External supply chain complexity, Supply chain performance, Upstream supply chain complexity, Operational supply chain complexity, Downstream supply chain complexity. Similarly, Malesios *et al.* (2020) use the formative model Supply Chain Sustainability, following the previous papers by De Giovanni and Vinzi (2012) and De Giovanni (2012) exhorting the use of formative models for the whole area of sustainability in operations management.

Unlike traditional SEM applications that rely on reflective indicators, Xu *et al*. (2019) argue for the importance of formative constructs in measuring complex constructs like quality management and supply chain performance. They provide empirical evidence supporting the idea that formative models can better capture cause-effect relationships in operations management. The study by Wallenburg and Weber (2005) proposes an empirical analysis using formative measurement models in SEM, which is defined as to be crucial for supply chain research by the authors.

Accordingly, the use of formative measurement models allow one to demonstrate that logistics service quality (timeliness, flexibility, and reliability) contributes more to financial and strategic performance than cost-cutting measures. According to Garver (2019), the misuse of reflective and formative measurement models in SEM leads to the improper construct specification and misleading conclusions. For example, if "supply chain agility" is driven by indicators like speed, flexibility, and adaptability, treating it as reflective distorts its theoretical foundation.

Park *et al*. (2023) examine how supply chain agility impacts firm performance while defining agility as an organization's ability to sense changes in the market and respond effectively. They use formative models even for the second-order factor constructs. Considering the various use of both formative and reflective models, the next section introduces the differences between the two measurement scales.

3. Differences between Reflective and Formative Scales

Table 2 summarizes the technical and conceptual differences between reflective and formative measures. It intends to drive researchers in O&SCM during the selection of the measurement model to be selected.

Table 2 – Differences between reflective and formative models

	Reflective	**Formative**
Direction of causality	From the construct to the measures	From the measures to the construct
Correlation among items	Measures expected to be correlated. Measures should possess internal consistency reliability	Measures expected not to be correlated. Internal consistency is not implied
Importance of items	Dropping an item from the scale does not change the meaning of the construct	Dropping an item from the scale may imply a change in the meaning of the construct
Measurement error	Accounted at the variable level	Accounted at the construct level
Interchangeability of items	Items are interchangeable	Items are not interchangeable
Covariation among items	Indicators are expected to covary with each other	Indicators are expected not to necessarily co-vary with each other
Nomological net (indicators have the same antecedences and consequences)	Should not differ	Differ
Multicollinearity	Required	It is a serious concern

Author's development

The main assumption behind the usage of reflective measurement models consists of having a set of measures whose covariance is caused by a variation in latent variable. The causality relationship goes from the construct to the variables, thus the model is reflective because the constructs reflects the manifest variables (Fornell and Bookstein 1982). In contrast, a formative model assumes that all measures have an impact on – and thus cause – the construct and jointly determine its meaning (Bollen and Lennox, 1991; Fornell and Bookstein, 1982). Causality goes from the indicators to the construct, measures do not covary and are not at all correlated. Therefore, internal consistency does apply for reflective models while it does not for formative ones.

The concern around the selection of a measurement model in O&SCM research emerges from the first steps of an empirical research project. According to the selected mode, the researcher will define the wording of each question. Questions will be stated in a passive form when the measurement model is reflective, and in an active form for a formative measurement model. For instance, to measure the level of integration with suppliers, De Giovanni (2020) used (among others) the measure *sharing demand forecasting* whose information has been obtained through the question:

«We share our demand forecasting with our major suppliers» (De Giovanni, 2020).

Should the causality relationship go from the construct (integration with suppliers) to the measure (sharing demand forecasting) or *vice versa*? While the author used a reflective measurement model, researchers should devote more attention to the sense and meaning of the question to correctly obtain the needed information. The following doubts emerge:

Is a company integrated with suppliers because it shares demand forecasting?

or

Because a company is integrated with suppliers then it shares demand forecasting?

In the former case, a formative model should be selected, in the latter case a reflective model is the most appropriate. The problem does not only rely on the way of wording the question, but also on how it is perceived from the interviewed person. The answer should be given according asking "*who comes for first? Integration with suppliers or sharing demand forecasting?*". Indeed, leaving this lack of details increases the biasness on the obtained information and the accuracy of derived results.

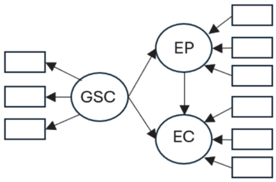
Along with the causality direction, also the number of items to be included in the scale assumes a peculiar importance. Missing one or more indicators under a formative mode can be a serious concern because an important part of a construct can be disregarded and its meaning can be completely different. A census of all possible measures to include in the scale should be carried out before start the data collection (Bollen and Lennox, 1991) while the elimination and the inclusion of a measure must be theoretically justified. Because each indicator contributes to create (form) the construct, measures are not expected to covary but rather to be independent dimensions. The standard scale development procedures (for example, dropping items that possess low inter-item correlations) does not apply in formative models as some items can be dropped, although they provide an extremely important contribution in the formation of a construct. This confirms the inappropriateness of internal consistency reliability for formative models and leads indicators not to be interchangeable.

In a formative measurement model the errors are associated with the construct rather than with the items while, in contrast, errors are associated with the items rather than with the construct in reflective scales. The error term supplies information on the appropriateness of the selected measures. According to MacKenzie *et al*. (2005, 712) the error term captures the invalidity of the set of measures caused by measurement error, interactions among the measures, and/or aspects of the construct domain not represented by the measures. Diamantopoulos (2006) demonstrates that the error term does not represent measurement error because formative indicators are specified to be error-free; rather, it represents the impact of all remaining causes other than those represented by the indicators included in the model. Therefore, it results that a reflective treatment of a formative construct reduces the variance of the construct because the variance of a reflective construct equals the common variance of its measures, while the variance of a formative construct includes the total variance of its measures. Consequently, if a misspecification reduces the variance of the exogenous variable while the level of the variance of the endogenous variable is maintained, the parameter estimate for their relationship increases. In contrast, if a misspecification reduces the variance of the endogenous variable while the variance of the exogenous variable is unchanged, the relevant structural parameter estimate decreases. Finally, model misspecification leads to an over or to an underestimation of structural parameters, which brings undesirable effects on the interpretation of findings.

In contrast with reflective models, multicollinearity among indicators can be a significant problem in formative scales, therefore high inter-item correlations may imply a drop off of items. See Bollen and Lennox (1991), Diamantopoulos and Winklhofer (2001), and Diamantopoulos and Siguaw (2006) to check how multicollinearity can be handled. Errors during the scale development and purification leads to high parameter bias because omitting an indicator in formative mode can lead to a completely different construct. Under this perspective, knowing that this analysis only leads to indicator elimination on purely statistical basis and given the possible alteration in the meaning of a construct, Diamantopoulos and Winklhofer (2001) suggest that indicator «elimination should never be divorced from conceptual considerations when a formative measurement model is involved» (Diamantopoulos and Winklhofer, 2001, 273).

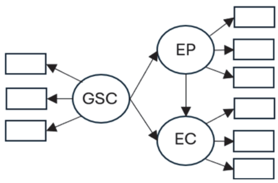
Finally, the vast majority of models incorporating misspecified measurement models show high acceptable values for chi-square per degree of freedom, CFI, GFI, SRMR, and RMSEA (Diamantopulos *et al.*, 2008). These indexes are not suitable to detect the correctness of a measurement scale thus a good fit should not mislead researchers from a correct model specification.

Figure 2 – Various combinations of models and errors



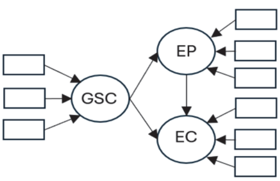
Model 1. Correct model specification

Research hypothesis	Standardized coefficient	p-value	Result
GSC ->EP	0.516	<0.05	Supported
GSC ->EC	0.283	<0.05	Supported
EP -> EC	0.473	<0.05	Supported



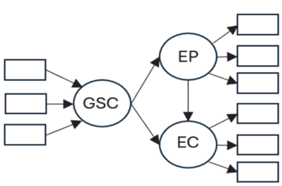
Model 2. Error of Type I on Performance

Research hypothesis	Standardized coefficient	p-value	Result
GSC ->EP	0.588	<0.05	Supported
GSC ->EC	0.596	<0.05	Supported
EP -> EC	0.212	>0.05	Non-Supported



Model 3. Error of Type II on GSC

Research hypothesis	Standardized coefficient	p-value	Result
GSC ->EP	0.554	<0.05	Supported
GSC ->EC	0.285	<0.05	Supported
EP -> EC	0.460	<0.05	Supported



Model 4. Error of Type II on GSC and Type I on Performance

Research hypothesis	Standardized coefficient	p-value	Result
GSC ->EP	0.470	>0.05	Non-Supported
GSC ->EC	0.230	>0.05	Non-Supported
EP -> EC	0.473	<0.05	Supported

Author's development

4. An Example of Model Misspecification in O&SCM

Hereby, I provide an example to demonstrate the changes in the empirical results according to the measurement models selected. Specifically, I analyze the impact of *environmental collaboration with suppliers* (Green SC, GSC) on Environmental Performance (EP) and economic performance (EC). Please, check De Giovanni and Cariola (2020) for a complete overview on the research hypotheses and scale development. For a correct identification of the model, GSC should be measured through a reflective scale, while EP and EC should be measured through a formative measurement model. I have taken the data displayed in the correlation table in De Giovanni and Vinzi

(2012) to run the empirical analysis and to develop the four cases reported in Figure 2.

This example supplies information on the possible consequences due to an inappropriate selection of the measurement models. Specifically:

a) *Model 1*. The exogenous construct GSC is correctly specified through a reflective model; both the constructs of performance are correctly specified through formative scales. Hence, the model is correctly specified and the results of the hypotheses are reliable;
b) *Model 2* highlights misspecification in the endogenous constructs linked to Performance, which are modelled in a reflective mode rather than in a formative mode. This generates the hypothesis EP→EC to be non-significant, while it results to be significant when correctly specified as in Model 1;
c) *Model 3* has misspecification in the exogenous construct linked to GSC. This leads to an interpretation error of the meaning of the constructs and the effects; for example, since all constructs are reflective, GSC influences positively the firms' capacity to increase the economic and the environmental performance. Instead, the interpretation of the Model 1 is different: GSC influences positively the firms' economic and environmental performance. In sum, Model 3 insists on the firms' "capacity" to perform while Model 1 works on the firms' "performance" itself. Therefore, the use of different models, being both highly significant, impacts importantly on their meaning and interpretation;
d) *Model 4* presents misspecification in both the exogenous construct of GSC and the endogenous constructs of Performance. Notice that GSC→EP and GSC→EC have different results with respect to Model 1; hence, the findings obtained are not reliable.

The distinction between reflective and formative measurement models in SEM is rooted in their foundational assumptions about causality, indicator relationships, and statistical methodologies for estimation and validation. These models necessitate distinct approaches to handling data distribution, estimation methods, and the application of specific fit indices for model assessment.

Reflective models are very popular since they conceptualize latent constructs as underlying causes that manifest through observed indicators. The relationship is mathematically represented in matrix form as $X = \Lambda \xi + \varepsilon$, where X represents the matrix of observed indicators, Λ is the matrix of loadings that link the latent variables ξ to the indicators, and ε is the vector of normally distributed error terms for each indicator. This model structure leads to the covariance matrix of X being modeled as $\Sigma = \Lambda\Phi\Lambda$^T + Θ, where Φ is the covariance matrix of the latent variables, and Θ is the diagonal matrix containing variances of the error terms. Estimation techniques such as Maximum Likelihood (ML) or Generalized Least Squares (GLS) are used, assuming multivariate normality of the indicators, which can be validated using tests such as Kolmogorov-Smirnov or Shapiro-Wilk. Model evaluation for reflective models employs fit indices including the Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), alongside residual-based measures such as the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). These indices measure how well the proposed model's covariance matrix (Σ) reproduces the observed data's covariance matrix. Internal consistency is checked using Cronbach's Alpha and Composite Reliability (CR), ensuring the indicators reliably reflect the latent construct.

In contrast, formative models define latent constructs as composites formed by their indicators, expressed in vector form as $\xi = \pi$^T $X + \delta$, where π represents the row vector of weights assigned to each indicator in X, and δ is the error term at the construct level. This structure does not assume normal distribution of errors, reflecting the non-causal nature of the relationship between indicators and the construct. Formative models typically use Partial Least Squares (PLS), a method focusing on maximizing explained variance in dependent constructs without requiring distributional assumptions, suitable for complex model estimations where traditional covariance-based methods might fail. Multicollinearity among indicators is a critical consideration in formative models, assessed using the Variance Inflation Factor (VIF). Unlike reflective models, formative models do not utilize traditional goodness-of-fit indices due to the absence of a latent variable causing the observed indicators. Instead, model validation focuses on the significance and relevance of the weights (π) calculated for each indicator, often evaluated using bootstrap techniques to provide non-parametric confidence intervals.

Indeed, differently from reflective models, the formative models generate a certain amount of technical issues: these do probably suggest to the O&SCM to go for reflective options since there are no so many technical challenges. Beyond the t-rule and the scaling rule conditions requested by reflective models (Diamantopulos *et al.*, 2008), the rule requiring at least two emitted paths must also apply (Bollen and Davis, 1994). It consists on having two leaving arrows from the formative model, which allow the identification of the disturbance term. The two arrows may go either to other measures or to other constructs (Bollen and Davis, 1994) or to both (Jöreskog and Goldberger, 1975). As for the multicollinearity, adding arrows only for identification reasons puts the model specification into question if these outcomes are not theoretically supported. Instead of struggling with the inclusion of improbable research hypotheses, a different estimation method may substantially help. For instance, being a component-based estimation method, the Partial-Least Square (PLS) represents a valid alternative to estimate empirical models while eliminating the technical restrictions imposed by Lisrel, which is instead a covariance-based estimation method (CVA). The approach uses an iterative combination of principal components analysis and regression to explain the variance of each construct. Because PLS makes no distributional assumptions, traditional parametric significance testing procedures are not appropriate.

Table 3 – Comparison of PLS and CVA

Criterion	**PLS**	**CVA**
Objective	Prediction oriented	Parametric oriented
Approach	Variance based	Covariance based
Assumptions	Nonparametric	Parametric
Parameter estimates	Consistent at large	Consistent
Latent variable scores	Explicitly estimated	Indeterminate
Model complexity	High	Small to moderate
Minimum sample size	20-100	200-800

Author's development

Table 3 reports the key differences among component- (PLS) and covariance-based (CVA) approaches. PLS seeks to maximize prediction in the endogenous constructs rather than estimating covariances among latent variables. It is generally used for predicting behavior, with the final purpose to explain the model variance. PLS substantially helps when models are really complex to estimate as well as sample sizes are really small. The algorithm converges in a few iterations independently of measuring constructs through reflective or formative scales. A small sample size does imply any identification problem while large size increases the consistency of PLS estimations. Formally, PLS should be used any time requirements for multinormality, large sample size, and good model specification cannot be met. Due to PLS bias, structural estimations are understimaned while measurement model relationships are overstimated. CVA should be used when the goal is theory testing, theory confirmation, or the comparison of alternative theories, errors require additional specification (e.g., covariation), and the research requires a global goodness-of-fit criterion.

5. Conclusions

This research aims to contribute to a major and common error among researchers in the field of Operations and Supply Chain Management (O&SCM) and related to choosing the right measurement models in accordance with Structural Equation Modeling (SEM). A substandard or incorrect choice of a measurement model according in SEM may lead to misinterpretation of the findings, and erroneous conclusions drawn from the research. This is the very first-time research in O&SCM has elaborated on the differences between reflective and formative measurement models used in SEM, as well as notifying the researchers in this area and making them conscious and careful about the construction and selection of the measurement model before they embark in statistical analysis. Therefore, this research makes an original contribution to the body of knowledge in the O&SCM area and invitation to growth in the right direction.

5.1 Insights for firms and practitioners

The results of this research suggest that enterprises and professionals of O&SCM face a challenge when selecting a model to analyze a business phenomenon: the appropriate incorporation and application of measurement models under the SEM framework can deeply impact the preciseness of the strategic decisions that are made based on analytics. For example, employing a formative model for production efficiency, which is a composite indicator derived from several operational metrics (e.g., material cost, scrap rate, and energy consumption), ensures that the decision-makers do not overlook vital dimensions of performance. Additionally, the use of reflective models might enable in the better understand of the underlying latent constructs such as organizational culture or customer satisfaction which are reflected across various observable indicators. Therefore, firms can benefit greatly by using these models through the right methods for making their strategies and work by correctly interpreting the data.

5.2 Insights for research and academia

This research reveals an important evidence in the academic literature in O&SCM with respect to the right SEM measurement models selected by researchers. Therefore, this paper challenges the academic community to improve their methodological rigor not only by choosing the right measurement models but also by having a clear understanding of the theoretical basis that justifies their choice. Researchers are advised to look into SEM intricacies, which are a basis for more reliable and more credible research findings. The study also indicates the need for educational curricula to include the construction and analysis of advanced statistical techniques, given the close relationship between these decisions and research outcomes. In fact, the selection of the measurement model in O&SCM research should be driven by theoretical foundation rather than being thoughtlessly carried out through reflective scales.

This would lead research to a correct model specification, that is, to a truthful establishment of the nature and direction of relationships between constructs and measures. The consequences of measurement model misspecification consist of serious under- or overestimations of parameters due to misidentified causality, wrong purification procedures, or a combination of both. Such biases may in turn lead to incorrect conclusions on tested hypotheses. This is especially true in the light of the fact that a satisfactory overall model fit does not guarantee a correct specification while misspecifications are not detected by poor fit index values. Hence, research in O&SCM should devote more attention to the measurement model selection to final deliver valid contributions to the literature and truthful findings.

Acknowledgement

I would deeply thank Prof. Salvatore Esposito De Falco for editing this special issue in memory of Prof. Alfio Cariola. I would also thank the Editor and the three anonymous Reviewers for the precious feedback and the opportunity to improve the manuscript in the various review rounds. This research was undertaken at the DIR – Claudio Demattè Research Division, within the Sustainable Operations and Supply Chain Monitor at SDA Bocconi School of Management (Milan, Italy). Special thanks go to all the companies that supported the research project. In alphabetical order, our thanks to: Carpe Diem Valuenet S.r.l., Celebron S.r.l., DNV Business Assurance Italy S.r.l., Edison Energia S.p.A. and Edison Next S.p.A., Elex Italia S.c.P.A., Evoca S.p.A., EY S.p.A., FERCAM S.p.A., Fincantieri S.p.A., Merck Serono S.p.A. (an affiliate of Merck KGaA, Germany), Kontractor by Kopron S.p.A., Mundys S.p.A., Nesea-NSA Italia S.r.l., Prologis Italy Management S.r.l., Salov S.p.A. (Filippo Berio and Sagra brands), Spinosi Marketing Strategies S.r.l. and Verity AG.

References

Alsheyadi A., Baawain A., Shaukat M.R. (2024). E-supply chain coordination and performance impacts: An empirical investigation. *Production & Manufacturing Research*, 12(1): 2379942. DOI: 10.1080/21693277.2024.2379942

Bollen K. and Davis W. (1994). *Causal indicator models: identification, estimation, and testing*, Paper presented at the American Sociological Association Convention, Miami.

Bollen K., Lennox R. (1991). Conventional wisdom on measurement: a structural equation perspective, *Psychological Bulletin*, 110(2): 305–14.

Bollen K.A. (1989), *Structural Equation with Latent Variables*, Wiley, New York, NY.

Chand P., Kumar A., Thakkar J., Ghosh K.K. (2022). Direct and mediation effect of supply chain complexity drivers on supply chain performance: an empirical evidence of organizational complexity theory. *International Journal of Operations & Production Management*, 42(6): 797-825.

De Giovanni P. (2012). Do internal and external environmental management contribute to the triple bottom line? *International Journal of Operations & Production Management*, 32(3): 265-290.

De Giovanni P., Cariola A. (2021). Process innovation through industry 4.0 technologies, lean practices and green supply chains. *Research in Transportation Economics*, 100869. DOI: 10.1016/j.retrec.2020.100869

De Giovanni P., Vinzi V.E. (2012). Covariance versus component-based estimations of performance in green supply chain management. *International Journal of Production Economics*, 135(2): 907-916. DOI: 10.1016/j.ijpe.2011.11.001

De Giovanni, P. (2020). When feature-based production capabilities challenge operations: Drivers, moderators, and performance. *International Journal of Operations & Production Management*, 40(2): 221-242. DOI: 10.1108/IJOPM-04-2019-0309

Diamantopoulos A., Siguaw J. (2006). Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration, *British Journal of Management*, 17(4): 263–82.DOI: 10.1111/j.1467-8551.2006.00500.x

Diamantopoulos A., Winklhofer H. (2001). Index construction with formative indicators: an alternative to scale development, *Journal of Marketing Research*, 38(2): 269–77. DOI: 10.1509/jmkr.38.2.269.18845

Fornell C., Bookstein F.L. (1982). A comparative analysis of two structural equation models: LISREL and PLS applied to market data. In Fornell C. (Ed.). *A second generation of multivariate analysis*, vol. 1. New York: Praeger, pp. 289–324.

Garver M.S. (2019). Threats to the validity of logistics and supply chain management research, *Journal of Business Logistics*, 40(1): 30-43. DOI: 10.1111/jbl.12203

Jarvis C.B., Mackenzie S.B., Podsakoff P.M., Mick D.G., Bearden W.O. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research, *Journal of Consumer Research*, 30(2): 199-218. DOI: 10.1086/376806

Jöreskog K., Goldberger A. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable, *Journal of American Statistics Association*, 10: 631–9. DOI: 10.1080/01621459.1975.10482485

MacKenzie S., Podsakoff P., Jarvis C. (2005). The problem of measurement model misspecification in behavioural and organizational research and some recommended solutions, *Journal of Applied Psychology*, 90(4): 710–30. DOI: 10.1037/0021-9010.90.4.710

Malesios C., Dey P.K., Abdelaziz F.B. (2020). Supply chain sustainability

performance measurement of small and medium sized enterprises using structural equation modeling. *Annals of Operations Research*, 294(1): 623-653.
Petter S., Straub D., Rai A. (2007). Specifying formative construct in information system research, *MIS Quarterly*, 31(4): 623-656. DOI: 10.2307/25148814
Podsakoff P.M., MacKenzie S.B., Podsakoff N.P., Lee J.Y. (2003). The mismeasure of man (agement) and its implications for leadership research. *Leadership Quarterly*, 14: 615-656. DOI: 10.1016/j.leaqua.2003.08.002
Roh T., Noh J., Oh Y., Park K.S. (2022). Structural relationships of a firm's green strategies for environmental performance: The roles of green supply chain management and green marketing innovation. *Journal of cleaner production*, 356: 131877. DOI: 10.1016/j.jclepro.2022.131877
Wallenburg C.M., Weber J. (2005). Structural equation modeling as a basis for theory development within logistics and supply chain management research. In Kotzab E., Seuring S., Müller M., Reiner G. (Eds.), *Research methodologies in supply chain management*, Springer, 171-186.
Xu L., Peng X., Prybutok V. (2019). Formative measurements in operations management research: Using partial least squares. *Quality Management Journal*, 26(1): 18-31. DOI: 10.1080/10686967.2018.1542287