IPAM THE MARKETING SCHOOL

SOME NEW THOUGHTS ON FORMATIVE AND REFLECTIVE MEASUREMENT MODELS IN MARKETING RESEARCH

KEYWORDS: FORMATIVE MODELS, MEASUREMENT, MODELLING, MARKETING

ALGUNS NOVOS PENSAMENTOS EM MODELOS FORMATIVOS E REFLEXIVOS NA INVESTIGAÇÃO DE MARKETING

PALAVRAS-CHAVE: MODELOS FORMATIVOS, MEDIÇÃO, MODELAÇÃO, MARKETING

JOSE A. MARTÍNEZ
Universidad Politécnica de Cartagena

LAURA MARTÍNEZ
Universidad Politécnica de Cartagena

Abstract

We discuss the current conceptualization of reflective and formative measurement models in marketing research. We reformulate the definition of formative models as follows: in a formative measurement model in which a construct is defined in terms of its measurements, the variable of interest susceptible to be measured is an algebraic construction that cannot be measured in a reflective way. We also introduce a new type of indicator, the phantom-effect indicator, which is a numerical trick used to correctly identify some types of causal models. We based our thinking on the definition of a latent variable and the interpretation of the error term, breaking from the current thinking that formative variables do not exist independently from their indicators. In addition, we propose that variables can be classified into three disparate categories with regard to the form of measurement: cake, love, and player performance. We use these terms to illustrate our reasoning. Finally, we offer guidelines to model formative variables when they are causes or consequences of other variables.

Resumo

O artigo discute a actual conceptualização de modelos formativos e reflexivos na investigação de marketing. Nós reformulamos a definição de modelos formativos como se segue: num modelo de medição formativo em que um constructo é definido em termos das suas medições, a variável de interesse susceptível de ser medida é uma construção algébrica que não pode ser medida de uma maneira reflexiva. Também introduzimos um novo tipo de indicador, o efeito de phantom, que é uma solução numérica utilizada para identificar correctamente alguns tipos de modelos causais. Nós baseamos o nosso pensamento na definição de uma variável latente e interpretação do termo de erro, afastando-nos do pensamento actual de que as variáveis formativas não existem independentemente dos seus indicadores. Adicionalmente, nós propomos que as variáveis possam ser classificadas em três categorias distintas relativamente á forma de medida: bolo, amor e jogador. Nós usamos estes termos para ilustrar os nossos argumentos. Finalmente, são oferecidas orientações para modelar variáveis formativas quando elas são causas ou consequências de outras variáveis.

ESPAÇO NOTAS

Introduction

Measurement is one of the most important concerns for marketing researchers. There are two approaches to measure a variable of interest: the reflective model and the formative model. The variable susceptible to be measured is a theoretical entity that we want to know, and we need to use a valid measurement instrument to know about it. Bollen (2002) and Edwards and Bagozzi (2000) explain the relationship between theory, models, and variables more succinctly.

A valid measurement instrument is composed of one or more indicators. Indicators are called observable variables and they provide values or scores. These values are close to the unknown values of the theoretical entity. If the measurement instrument is perfect, there is no difference between observable and theoretical values. However, in social sciences, there is some degree of imperfection in the measurement instruments. Measurement error determines the relationship between the theoretical entity and the observable indicators.

In a reflective model, variations in the theoretical entity are reflected in variations in its observable measures valid for this purpose. These types of indicators are called "effect indicators." However, in a formative model, the theoretical entity is defined in terms of its measurements. The theoretical entity is *composed of*, not *defined by*, observable indicators. Therefore, these types of indicators are "cause indicators." Relevant references in market research show how to distinguish between reflective and formative models and give recommendations for modelling these two different conceptualizations (Diamantopoulos and Winklhofer, 2001; Jarvis, MacKenzie, and Podsakoff, 2003). A more profound theoretical discussion can be found in Edwards and Bagozzi (2000) and Borsboom, Mellenbergh, and van Heerden (2003).

Jarvis, MacKenzie, and Podsakoff (2003) show that 28% of the theoretical entities measured with multiple indicators published in the top marketing journals were incorrectly specified as reflective when they should have been formative. Indeed, this type of measurement model misspecification affects several of the most commonly used variables in the field. In addition, they demonstrated that measurement model misspecification severely biases structural parameter estimates and can lead to inappropriate conclusions about the hypothesized relationships between the variables. Therefore, measurement relationships must be appropriately modelled. Diamantopoulos, Riefler, and Roth (2008) also comment about reviews that show how theoretical entities were incorrectly specified as reflective when they should have been formative.

The philosophical and practical aspects of reflective and formative models are still a matter of debate, see, e.g., the diverse perspectives of Bagozzi (2007), Bollen (2007), Howell, Breivik, and Wilcox (2007), and Coltman, Devinney, Midgley, and Venaik (2008). Our aim is not to discuss all the potential issues that are susceptible to be treated (e.g., Wilcox, Howell, and Breivik, 2008), neither to deepen into the philosophical debate regarding different approaches, but to introduce our position regarding the nature of formative variables and the empirical consequences to measure a variable using an algebraic construction of cause indicators or using effect indicators. We will show that the current consideration of formative variables is ambiguous, we distinguish formative variables from purely algebraic constructions, and we propose that variables can be classified into three categories cake, love, and player performance. The rationale of these names comes from the illustrative examples that we used to develop our reasoning. Only variables that fall into the "player performance" category should be considered as formative. In addition, we will demonstrate that "cake" and "love" variables can be measured using reflective indicators and using a composite of indicators, and that they are empirically equivalent in the mean statistic, though they differ in the variance if measurement error exists. Only our understanding of formative variables guarantees that they are not empirically equivalent to variables measured using reflective indicators. Finally, we introduce the concept of phantom-effect indicators, a necessary numerical trick to estimate some causal models. Therefore, this new understanding of formative variables is the major contribution of this research.

Theoretical entities as latent variables

We consider the theoretical entities as the variables susceptible to be measured as latent variables. The definition of latent variable has also been a matter of debate. Bollen (2002) discusses several definitions; from the common use of the term: "concepts that are not directly observable," to the more formal definitions based on the following principles: (1) local independence, (2) classical test theory-expected value, (3) nondeterministic function of observed variables, and (4) sample realization. Bollen (2002) argues that the last one is the most inclusive definition, which covers a broad range of methodologies and models. However, Bollen (2002) concludes that there is no right or wrong definition of latent variables, and it is more a question of finding the definition that is most useful depending on the context of application.

We focus our reasoning on the most used research framework in market research: when we are interested in the measure of variables such as perceived quality, market orientation, and future repurchase intentions, we are interested in knowing the descriptive statistics of these variables, how individuals are positioned within the domain of these variables, and establishing relationships between variables by conducting multivariable statistical analyses. Under these general frameworks, we consider latent variables following the principles of classical test theory. As Borsboom et al. (2003) explain, classical test theory views measurement in statistical terms. A score yielded by an observable indicator is a measure of a theoretical construct if its expected value increases monotonically with that construct. Therefore, the theoretical construct could be taken to be the true score. Measurement error affects the observable variance of the observable indicator but does not bias its expected value. This perspective is compatible with the realist view of science. In addition, realism is associated with causality. Hence, theoretical entities are causally responsible for observed phenomena, so the latent variable and indicators are distinct entities.

Bollen (2002, p. 613) implicititly assumes that the existence of measurement error distinguishes a latent from an observed variable, when he speaks about the thermometer example; contemporary thermomethers are still not perfect, so that thermometer readings are not synonymous with temperature. A similar view regarding latent variables is adopted by Hayduk (1996); measurement error determines the relationship between the theoretical entity and the observable indicators. Hayduk (1996) considers variables such as sex or age as latent variables, because measurement error could exist (imperfect recall, coding mistakes, etc.). Therefore, all variables susceptible to be measured are latent, because there is some degree of measurement error in their observable indicators. If the researcher considers that measurement error is zero or negligible in a given sample, then the scores of the observed variables perfectly reflect the scores of latent variables, but continue being distinct entities. Obviously, this is a realist interpretation.

We must stress that the true score of the latent variable can be computed by using the expected value of the observable indicators, provided there is no systematic bias in this indicator. For example, some indicators of theoretical entities such as "degree of xenophobia" could be susceptible to get affected by a systematic error related to social desirability bias. If this type of systematic error exists, then obviously there is a bias in the estimation of the expected value of the theoretical entity.

Hayduk (1996) emphasizes the necessity of a correct definition of the latent variable, because the meaning of the latent depends on the relationship between the abstract concept and the real-world manifestation. This is an *a priori* definition, and the researcher has to establish his/her commitment to the meaning of the latent through the correspondence to the latent and the best indicator or each theoretical entity. This is achieved by fixing the measurement error to a specific value.

However, a different view of measurement is possible, where the variable of interest is nothing more than its empirical content, a numerical trick used to simplify observations (Borsboom et al., 2003). These relationships are shown in a formative model in which a construct is defined in terms of its measurements. The variable of interest is a composite of indicators. Therefore, variations in the observed indicators cause variations in the variable of interest. From an operational view, a concept that is measured with formative indicators is merely an algebraic construction. There is no causal statement beyond the organizing principle of the representation (Markus, 2004), and there is no distinction between the construct and its measures because the construct is defined in terms of its measures. This is consistent with the constructivist view of science. Therefore, some authors argue that formative theoretical entities do not exist independent of measurement as reflective entities do (e.g., Diamantopoulos, 2006). We first comment a "candy" example to clarify this important issue.

Our favourite cake is composed of sugar, chocolate, and milk, but this cake is a different entity from sugar, chocolate, and milk. We know from Gestalt's philosophy and the science of complexity that the whole is more than the sum of its parts, both reductionism and determinism being very questionable paradigms. Obviously, the flavour, texture, appearance, and other characteristics of our favourite cake depend on the correct combination of the three ingredients, i.e. variations in the weight of the ingredients cause variations in the flavour and other characteristics of the cake. The flavour of the cake also depends on several factors, as the temperature of the oven or the quality of ingredients, and other elements related to the experience of eating it, such as being accompanied by the desired person. However, the flavour of the cake is a subjective entity distinct from other characteristics of the cake, for example, its weight. The weight of the cake is a composite of the weight of the ingredients, i.e. an algebraic construction, and clearly does exist separately from the weight of its ingredients. In fact, the weight of the cake can be directly measured by weighing the cake on a balance. So, a composite variable such as cake weight, derived from these three indicators (the weight of each of the three ingredients), is not a formative variable under the definition of Diamantopoulos (2006) and Borsboom et al. (2003). Note that those authors consider formative variables as inextricably tied to their measures, i.e. they do not exist independently of measurement. A corollary of this example is that the definition of a formative variable has to be reformulated as follows: *in a formative measurement model in which a construct is defined in terms of its measurements, the variable of interest is a composite of indicators. Therefore, variations in the observed indicators cause variations in the variable of interest. The variable of interest is an algebraic construction that cannot be measured in a reflective way.*

We propose another example to understand our view regarding theoretical entities that have to be measured by formative models, and consequently cannot be measured in a reflective way. This is the case of basketball player performance. Player performance is one of the key variables that determine the market value of a basketball player. Measuring this variable is extraordinarily important in the NBA (Hollinger, 2005; Oliver, 2004), and many of the teams pay many dollars to contract private statistical services to evaluate the performance of players.

Player performance is a composite variable derived from several indicators representing positive actions (e.g., points, rebounds, etc.) and negative actions (e.g., field goals missed, fouls, etc.) of the player in the court. Variations in the observed indicators cause variations in the player performance; it is an algebraic construction that cannot be measured in a reflective way. We could measure subjective player performance in a reflective way (for example, through an expert's judgment), but this would be a different variable from the more "objective" player performance variable derived from the positive and negative actions of a player in the court. Bollen (2007) reasons in a similar way when he distinguishes between objective socioeconomic status (SES) from subjective SES, in his reply to Howell et al.'s (2007) advocacy for avoiding formative indicators.

At this point, it seems evident that there are theoretical entities that can be measured by both using causal indicators and effect indicators, such as the weight of a cake. However, some theoretical entities can only be measured with cause indicators, such as "objective" player performance. Our new conception of formative models differs from the current conception, because we consider only those algebraic constructions as formative that cannot be measured in a reflective way. Nevertheless, to develop our reasoning, we will use the term "formative*" to refer to the previously known concept of formative models. The key concern of formative models is the consideration or not of the variable of interest as a latent variable. We will further state our position regarding this concern.

Beyond the discussion about the confronted realist versus constructionist perspectives (e.g., Hunt, 1991; 1992), we must remember Markus (1998)'s advocacy of constructionist positive contribution: imagining a world where everything comes with a label, something like a metaphysical supermarket. This world is the realist paradise. In such a world, we are trapped inside a fixed and inflexible lexicon where the language is sometimes described as a prison. As Markus asserts, our ability to adapt to emergent social issues stems from our ability to reconfigure the way in which we connect words to things. This capacity is derived from what constructivism brings to our understanding of language.

Therefore, the a priori definition of the variable we want to measure is absolutely vital. For example, How to measure love? There is a chemical reaction in the brain when a person is in love with another person (Punset, 2007). But how do we define love? This is the previous task to measure it. One researcher could define love as a feeling of attraction to the other person. A second researcher could define love as a feeling of care for another person. However, a third researcher could define love as a composite of two feelings: attraction and care. The first two researchers would be reasoning in a reflective way, and the third in a formative* way. Is there a consensus to define love? It seems clear that the definition of love depends on several factors such as culture or religion.

And what is the definition of player performance? Depending on the context, player performance is defined in several ways. For example, in the NBA, there are several distinct measures of player performance[1], such as "tendex," "player efficiency rating," or "adjusted plus/minus" (Hollinger, 2005). Euroleague and other European competitions use other measures, such as "ranking." There is no consensus among experts regarding which is the more proper measure of player performance. Therefore, player performance inevitably depends on its definition, as theoretical entity, i.e., it depends on the selection of the indicators (e.g., points, rebounds, etc.) and the weights of those indicators.

It is evident that this form of conception makes the measurement instrument extremely sensitive to the definition of the theoretical entity. One of the criticisms that some authors associate with formative* models is that all relevant indicators should be included in the model, that is, a census of components should be used (Bollen and Lennox, 1991; Diamantopoulos and Winklhofer, 2001). This means that to define theoretical entities such as player performance, we need to include all relevant facets that form the concept. However, authors such as Rossiter (2002) propose to relax this restrictive condition using a panel of experts to select the most relevant indicators. This is a common practice in other scientific disciplines, such as medicine. For example, Colditz, Atwood, Emmons, Monson, Willett, Trichopoulos, and Hunter (2000) used group consensus among researchers at the Harvard Medical School and Harvard School of Public Health to identify risk factors as definite, probable, and possible causes of cancer. Risk points were allocated according to the strength of the causal association and summed. They created the Harvard Cancer Risk Index, which offers a simple estimation of personal risk of cancer. This is a formative variable composed of several risk factors that act as formative indicators, and this variable cannot be measured in a reflective way.

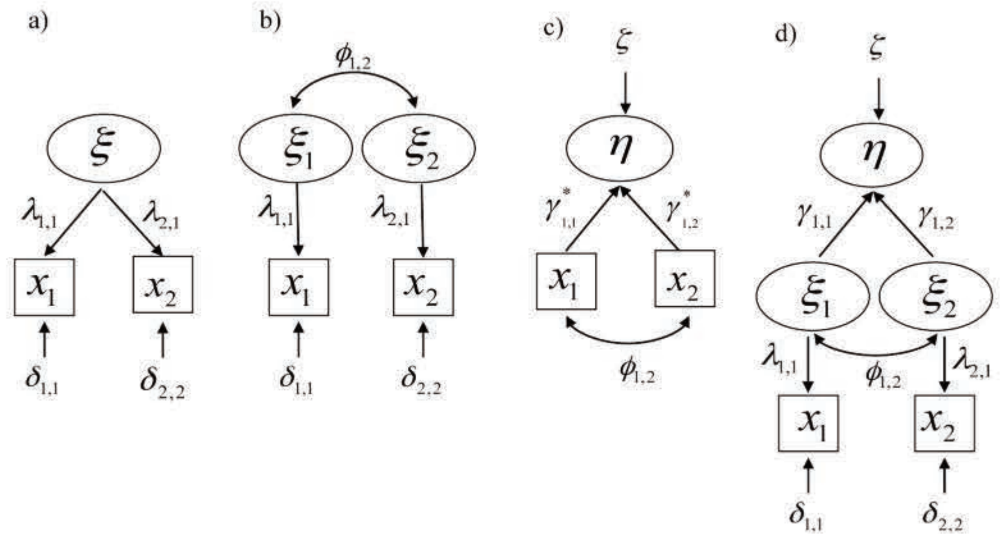
[1] A list of several measures of player performance can be consulted in: www.nbastuffer.com/component/option,com_glossary/Itemid,90/catid,42/func,display/

The specification of reflective and formative models

The specification of equations of reflective and formative* measurement models is well known (e.g., Bollen and Lennox, 1991). However, there are some types of measurement models, such as spurious models (Edwards and Bagozzi, 2000), that deserve special attention. Figure 1 illustrates four types of measurement models. We are going to depict the equations derived from these specifications to explain why formative* and reflective models are empirically equivalent in some statistics. We use the common structural equation modelling notation.

Figure 1

Four types of measurement models



Figures 1(a) and 1(b) represent two reflective models. In the first model, the theoretical entity is measured by two reflective indicators. In the second model, there are two distinct theoretical entities measured only with one indicator for each one. An example of the first specification is the measure of the flavor of a cake. One researcher could argue that the flavor of a cake can be defined by two indicators (a), such as "I like the taste of the cake," or "When I eat the cake I feel good." This type of specification must comply with the principle of local independence; that is, if a latent variable underlies a number of observed variables, then conditionally that latent variable will render the observed variables statistically independent. This condition means that equally reliable indicators of a latent variable are interchangeable. If the latent variable and indicators have the same scale, each single indicator should estimate the mean, and each indicator should not differ significantly in mean value from the rest. On the one hand, high correlations between indicators are not enough to represent the latent mean correctly, because if indicators are interchangeable, the selection of an indicator with a low score would distort the true latent mean.

The expressions for the mean and the variance of the variable of interest are (1):

$$\begin{aligned}
x_{1i} &= \lambda_{1,1}\xi_i + \delta_{1,1i} \\
x_{2i} &= \lambda_{2,1}\xi_i + \delta_{2,2i} \\
E(x_1) &= \lambda_{1,1}E(\xi) + E(\delta_{1,1}) \\
E(x_2) &= \lambda_{2,1}E(\xi) + E(\delta_{2,2}) \\
E(\xi) &= E(x_1)/\lambda_{1,1} = E(x_2)/\lambda_{2,1} \\
Var(x_1) &= \lambda_{1,1}^2 Var(\xi) + Var(\delta_{1,1}) \\
Var(x_2) &= \lambda_{2,1}^2 Var(\xi) + Var(\delta_{2,2}) \\
Var(\xi) &= \left[Var(x_1) - Var(\delta_{1,1})\right]/\lambda_{1,1}^2 = \left[Var(x_2) - Var(\delta_{2,2})\right]/\lambda_{2,2}^2
\end{aligned} \tag{1}$$

where x_{1i} are the responses of the *i* individuals to the first indicator, x_{2i} the responses of the *i* individuals to the second indicator; $\delta_{1,1i}$ and $\delta_{2,2i}$ are the measurement error for the two indicators, respectively; ξ_i is the value of the theoretical entity for the *i* individual; $\lambda_{1,1}$ and $\lambda_{2,1}$ are the coefficients giving the expected effects of ξ_i on x_{1i} and x_{2i}, respectively. We assume that x_{1i}, x_{2i}, and ξ_i are deviation scores around their means. ξ_i is uncorrelated with $\delta_{1,1i}$ and $\delta_{2,2i}$, cov($\delta_{1,1i}$, $\delta_{2,2i}$)=0 and $E(\delta_{1,1}) = E(\delta_{2,2}) = 0$. This is the classical specification shown in Bollen and Lennox (1991). In addition, if indicators and latent have the same scale, then $\lambda_{1,1}$ and $\lambda_{2,1}$ equals 1; this would be the case of tau-equivalent indicators (Jöreskog and Sörbom, 2001). If indicators are equally reliable, then $Var(\delta_{1,1}) = Var(\delta_{2,2})$; this would be the case of parallel indicators (Jöreskog and Sörbom, 2001). The latter is the most desired case for measurement, and as can be viewed in equation (1); this means that $E(\xi) = E(x_1) = E(x_2)$ and $Var(\xi) = \left[Var(x_1) - Var(\delta_{1,1})\right] = \left[Var(x_2) - Var(\delta_{2,2})\right]$, as we have explained earlier.

This first specification would also be the case when a researcher defines love through two indicators such as: "I am irremissibly attracted by this person" or "I care deeply about this person," or when a researcher defines subjective player performance such as "this player has played a great offensive game" or "this player has played a great defensive game."

However, another researcher could argue that the taste of the cake and feeling good when eating the cake are distinct theoretical entities. The same would occur if feeling of attraction and feeling of care are considered as different constructs. And if playing a great offensive game and playing a great defensive game are disparate things, as is the case in Figure 1(b), then the equations, under the same assumptions as before, would be (2)

$$\begin{aligned}
E(\xi_1) &= E(x_1) \\
E(\xi_2) &= E(x_2) \\
Var(\xi_1) &= \left[Var(x_1) - Var(\delta_{1,1})\right] \\
Var(\xi_2) &= \left[Var(x_2) - Var(\delta_{2,2})\right]
\end{aligned} \tag{2}$$

In this case, we continue with the assumption that x_{1i} and x_{2i} have the same scale of their respective latents. As it can be viewed, the expected values of the two latents do not have to be same. Both indicators do not have to be necessarily correlated, consequently the covariance of the latents ($\phi_{1,2}$) may take any value.

We consider a third option for measurement. This is the third case in Figure 1(c), when the variable of interest is a formative* variable composed of two formative* indicators. Consequently, the flavour of a cake is a composite of taste and feeling good; being in love is a composite of feeling attraction and feeling care, and subjective player performance is a composite of offensive and defensive performances. Formative* indicators do not have to be necessarily correlated. Furthermore, the flavour of a cake will increase to the extent that taste and feeling good increase, so the flavour of a cake will be higher when both indicators score high than when only one of them scores high. The same occurs for love and performance. Note the great difference of this reasoning with respect to the reasoning depicted in Figure 1(a), and note the similarities with the rationale shown in Figure 1(b)

The expressions for the mean and the variance of the variable of interest are (3):

$$\begin{aligned} \eta_i &= \gamma^*_{1,1}x_{1i} + \gamma^*_{1,2}x_{2i} + \zeta_i \\ E(\eta) &= \gamma^*_{1,1}E(x_1) + \gamma^*_{1,2}E(x_2) + E(\zeta) \\ Var(\eta) &= (\gamma^*_{1,1})^2 Var(x_1) + (\gamma^*_{1,2})^2 Var(x_2) + \gamma^*_{1,1}\gamma^*_{1,2} 2Cov(x_1, x_2) + Var(\zeta) \end{aligned} \quad (3)$$

where η_i is the variable of interest or theoretical entity, $\gamma^*_{1,1}$ and $\gamma^*_{1,2}$ are the coefficients giving the expected effect of x_1 and x_2 on η, and ζ is a disturbance term that represents the impact of all the remaining causes other than those represented by the indicators included in the model (Diamantopoulos, 2006). There is no measurement error in the indicators.

At this point, two important issues need to be clarified: the consideration of the variable of interest as latent in a formative* model and the interpretation of the disturbance term. We are going to comment on the last model before addressing these relevant concerns.

Figure 1(d) illustrates a re-specification of the third model. This is a formative* model with a reflective latent re-specification. These kinds of models are called spurious models (Diamantopoulos, 2006; Edwards and Bagozzi, 2000) and they are re-specifications of the formative* model that permits a realist interpretation. Therefore, the variable of interest is a composite of two latent variables that are measured in a reflective way. This would permit accounting for the measurement error in the observable indicators of each latent.

The expressions for the mean and the variance of the variable of interest are (4):

$$\begin{aligned} \eta_i &= \gamma_{1,1}x_{1i} + \gamma_{1,2}x_{2i} + \zeta_i \\ E(\eta) &= \gamma_{1,1}E(x_1) + \gamma_{1,2}E(x_2) + E(\zeta) \\ Var(\eta) &= (\gamma_{1,1})^2 \left[Var(x_1) - Var(\delta_{1,1})\right] + (\gamma_{1,2})^2 \left[Var(x_2) - Var(\delta_{2,2})\right] + \gamma_{1,1}\gamma_{1,2} 2Cov(x_1, x_2) + Var(\zeta) \end{aligned} \quad (4)$$

In this case, $\gamma_{1,1}$ and $\gamma_{1,2}$ are the coefficients giving the expected effect of ξ_1 and ξ_2 on η, and they have the same interpretation as of $\gamma^*_{1,1}$ and $\gamma^*_{1,2}$. However, we have distinguished the notation, because in the former case the coefficients relate to two observed indicators with the variable of interest and, in the latter case, to two latent variables with the variable of interest. The consideration of measurement errors adds value to this fourth model, and as can be derived for the expression (4), measurement error in the indictors affects.

The formative theoretical entity as a latent variable and the specification of the disturbance

Edwards and Bagozzi (2000) indicate that when the relationship of one measure with another is generally limited to deterministic functional relationships, as when one score is mathematically transformed into another score, or multiple scores are summed to create scales or item parcels, then these relationships are not causal; rather they simply entail mathematical operations to inert data. Thus, under this reasoning, an algebraic construction that it is a simple composite of observable measures is not a distinct variable from its indicators. However, we have shown how this assertion is very questionable, when we explained the example of the weight of the cake.

The reasoning of Edwards and Bagozzi (2000) is similar to the view of Diamantopoulos (2006) regarding the surplus meaning of the formative* constructs. Diamantopoulos (2006) comments that when the formative* measurement model does not include an error term, no surplus meaning can be attributed to the formative* construct; the latter simply becomes a weighted linear combination of its indicators. Under this view, the formative* variable is not a latent variable, because it does not exist separately from its indicators. Therefore, the specification of a nonzero disturbance (or error term) in the formative* variable is a criteria to decide if the formative* variable is latent or not, and whether it exists separately from its indicators or not.

Our position regarding this important issue is different. We consider formative variables as algebraic constructions that cannot be measured in a reflective way. The meaning of the formative theoretical entity can be distinct from its indicators, although their meaning depends on the combination of its indicators. Obviously, the variable "objective" player performance has a distinct meaning from points or rebounds, and in a similar form, SES has a distinct meaning from income or education.

The spurious model shows that indicators in a formative measurement model can be re-specified as latent variables. Recall that we consider latent variables as theoretical entities, which can be measured with observable indicators with some degree of measurement error. Theoretical entities are distinct from its indicators, but they are defined in terms of their indicators. A combination of theoretical entities is a combination of latent variables. Consequently, a formative variable is also a latent variable. Therefore, the specification of a nonzero disturbance term in a formative variable, as Diamantopoulos (2006) proposes, is not necessary to consider this variable as latent. Therefore, a formative variable is a composite of several latent variables, which can be measured with one or more indicators, and with a certain degree of measurement error. Measurement error of the indicators influences the variance of the formative variable. Here, we agree with Diamantopoulos (2006), because the disturbance or error term of the formative variable is not measurement error. Measurement error exists independent of the size of specification of the disturbance, as equation (4) indicates.

However, we disagree with Bollen and Lennox (1991), Edwards and Bagozzi (2000), Diamantopoulos (2006), and Bollen (2007) regarding the role played by the disturbance. Bollen (2007, p. 220), for example, considers that a formative variable is latent if there is a nonzero disturbance. But, the existence of the disturbance means that there are some other causes of the theoretical entity that have not been taken into account in the formative measurement model (Diamantopoulos, 2006). This implies that the theoretical entity is distinct from its indicators, or as spurious model shows, is distinct from the latent variables that form the variable of interest. Consequently, this implication matches with our aforementioned reasoning, i.e., formative variables are distinct entities from its indicators, but both reasoning are grounded on different premises.

The existence of the disturbance term is related with the surplus meaning of the construct. That is, the construct (or latent variable) contains meaning "over and above its simple and mathematical representation" (Podsakoff et al., 2003, p. 621). Diamantopoulos (2006) notes that any estimate of the error term—and hence the surplus meaning of the construct — is not only a function of the selected indicators but also depends on the selection of the additional constructs or measures used to attain model identification...and thus, as Wilcox, Howell, and Breivik (2008) claim, the selection of the "external" (i.e., additional variables necessary for achieving identification) is just as crucial in a formative* measurement model as is the selection of the formative* indicators themselves. Selection of the outcome variables is crucial but surplus meaning has nothing to do with the formatively measured construct. In the MIMIC model, the dependent variable (that which is regressed on the formative* indicators) is the shared variance of the reflected variables or constructs. Then, the error term is realized as the shared variance between the outcomes not accounted for by the formative* indicators. Thus, the meaning of the error term associated with formatively* measured constructs in structural equations models is more closely associated with the constructs dependent on the formative* construct and their correlation than on the formative* measures (Wilcox, Howell, and Breivik, 2008). Therefore, here there is the implicit assumption that a formative* theoretical entity has to be measured in a reflective way to identify its disturbance, and the size of this disturbance is highly dependent on the selected reflective indicators. This is what is called "interpretational confounding" (Burt, 1976). However, we do not consider theoretical entities as formative if they can be measured in a reflective way. Therefore, if there are some reflective indicators available to measure a variable, this variable, under our view, is not formative.

We see formative variables as theoretical entities that are defined in terms of its indicators. These indicators can be re-specified in a spurious model. Therefore, the formative theoretical entity is a latent variable, because it is a composite of latent variables, not because there is any disturbance. As disturbance is zero, equation (4) shows that the expected value and the variance of the formative latent variables can be determined from the observable indicators, making some assumptions regarding measurement error and the weights of these indicators. This is done when Hollinger (2005) calculates player performance through the formative variable "player efficiency rating." Hollinger assumes that measurement error is zero (there is no error in coding data, etc.) and he specifies the weights of the different indicators. Therefore, the expected value of the "player efficiency rating" of a player in a season can be easily derived from empirical data. The same occurs for the aforementioned example of the Harvard Cancer Risk Index, or in the case of the Apgar test (Apgar, 1953; Casey, McIntire and Leveno, 2001): a method of evaluation of the newborn infant.

The classification of marketing variables

We believe that the following classification of marketing variables can benefit researchers who share our viewpoint regarding the formative–reflective measurement debate. We classify variables as cake, love, and player performance. These are three categories that are subsequently explained:

A cake is referred to a physical entity that has properties. These properties can be physical (e.g., weight, density, height, etc.) or associated with subjective evaluations (e.g., taste, texture, appearance, etc.). Researchers can measure any of these properties by using the reflective approach, or by using an index (a composite of several indicators). This index is not a formative variable, under our definition, but an algebraic construction that is not fully empirically equivalent to the reflective measurement. The majority of marketing variables fall into this category: (e.g., perceived quality, perceived value, brand image, future repurchase intentions, market orientation, etc.). Variables such as age, sex, income, or ZIP code are also cake variables.

Love is referred to a mental state, a subjective perception of a feeling. Love is also referred to an attitude toward another physical or non-physical entity. It is similar to the subjective evaluations associated with the cake's category, but we distinguish one from the other because variables enclosed in the "love" category are more abstract, which cannot necessarily be linked to a physical object. For example, happiness is a theoretical entity that can be referred to an object (e.g., feeling happy while driving a car) or to a general feeling (e.g., life happiness). Other examples of these kinds of abstract variables would be depression, anxiety, optimism, etc.

Player performance is referred to algebraic constructions; variables that are derived from other variables and that cannot be measured in a reflective way. They are defined in terms of their indicators but have their own meaning. A variable that would fall into this category would be, for example, company performance, which could be measured as a composite of several indicators such as market share, sales, income, and return on investment. Another example would be the "classical" formative variable: the socioeconomic status, derived from income, education, occupational prestige, and neighbourhood. Obviously, as we have commented earlier, the meaning of the variables of this category depends on its definition, and a consensus regarding that definition would be desirable.

The empirical equivalence of cake and love variables

Cake and love variables can be measured using effect indicators (reflective models) or cause indicators (formative* models). They are empirically equivalent in some statistics. However, performance variables are not. We will use two examples of this equivalence and Figure 1 to illustrate our reasoning.

Example 1

To facilitate the demonstrations, we suppose that a cake or love variable is measured with only two cause (Figure 1d) or two effect indicators (Figure 1a). An example would be perceived service quality, which some authors treat as formative* (e.g., Rossiter, 2002) and others as reflective (e.g., Ko and Pastore, 2005). We consider perceived quality as a cake variable, because it is a subjective evaluation associated with a property of a service.

Descriptive statistics:

Mean and variance of the variable of interest in the case of effect indicators are:

$$\begin{aligned} E(\xi) &= E(x_1)/\lambda_{1,1} = E(x_2)/\lambda_{2,1} \\ Var(\xi) &= \left[Var(x_1) - Var(\delta_{1,1})\right]/\lambda_{1,1}^2 = \left[Var(x_2) - Var(\delta_{2,2})\right]/\lambda_{2,2}^2 \end{aligned} \quad (5)$$

Mean and variance of the variable of interest in the case of cause indicators are:

$$E(\eta) = \gamma_{1,1}E(x_1) + \gamma_{1,2}E(x_2)$$
$$Var(\eta) = (\gamma_{1,1})^2\left[Var(x_1) - Var(\delta_{1,1})\right] + (\gamma_{1,2})^2\left[Var(x_2) - Var(\delta_{2,2})\right] + \gamma_{1,1}\gamma_{1,2}2Cov(x_1, x_2) \quad (6)$$

We equate both equations $E(\xi) = E(\eta)$ and $Var(\xi) = Var(\eta)$, to analyze if these descriptive statistics depend on the specification of cause or effect indicators.

$$\gamma_{1,1}E(x_1) + \gamma_{1,2}E(x_2) = E(x_1)/\lambda_{1,1} = E(x_2)/\lambda_{2,1}$$
$$(\gamma_{1,1})^2\left[Var(x_1) - Var(\delta_{1,1})\right] + (\gamma_{1,2})^2\left[Var(x_2) - Var(\delta_{2,2})\right] + \gamma_{1,1}\gamma_{1,2}2Cov(x_1, x_2) =$$
$$= \left[Var(x_1) - Var(\delta_{1,1})\right]/\lambda_{1,1}^2 = \left[Var(x_2) - Var(\delta_{2,2})\right]/\lambda_{2,2}^2 \quad (7)$$

For the mean statistic, it is clear that if both x_1 and x_2 are measured in the same scale, $E(x_1)$ should equate $E(x_2)$, assuming that the scale of the theoretical entity is the same as its observable indicators. If both variables are measured using disparate scales, for example, a 5 and a 7 Likert-type scale, then both variables should be re-scaled to the unit [0,1] interval (Cohen, Cohen, Aiken, and West, 1999). Then, this is the common interpretation of reflective models under classical test theory: the mean of each one of the reflective indicators of a latent variable should estimate the latent mean.

Then, the only way to get $E(\xi) = E(\eta)$ is when $\gamma_{1,1}$ and $\gamma_{1,2}$ are weighted by 0.5. Therefore, the value of γ parameters should be divided by a number, which is the total number of indicators used to measure the variable of interest. For example, if three indicators are used, then γ parameters should be weighted by 0.33. Under these circumstances, the mean value of the variable of interest does not change when using effect or cause indicators. Obviously, this occurs if equally weighted γ are supposed.

However, the variance of the variable of interest depends on the measurement specification. As equation (7) illustrates, $Var(\xi)$ does not depend on the covariance of indicators, while this is not the case for $Var(\eta)$. In the case of equally reliable indicators, $\left[Var(x_1) - Var(\delta_{1,1})\right] = \left[Var(x_2) - Var(\delta_{2,2})\right] = a$, so the only way $Var(\xi) = Var(\eta)$ is when $Cov(x_1, x_2) = a$. When $Cov(x_1, x_2) \neq a$, then the variance of the theoretical entity is different when using cause than when using effect indicators.

Example 2

We consider now the measure of a physical characteristic such as the weight of a cake, produced by a factory. This type of variable is not common in marketing research, but we are going to use this example to show some divergences between the types of variables. Recall that in the case of the majority of marketing variables, the measurement scales are arbitrary, and the outcome of this measurement can be linearly transformed to a unit interval, i.e., the domain of the variable can be bounded to a unit interval. However, there are situations when the domain of a variable cannot be bounded in a closed interval, such as the weight, density, and height.

We consider a cake that is composed of two ingredients. We can measure the weight of the cake by using an instrument such as a balance. There are two options, to weigh the cake in the balance, or to weigh the three ingredients in the balance. In the former case, the weight of the cake provided by the balance is the effect indicator of the theoretical weight of the cake. In the latter, the weights of the two ingredients are the cause indicators of the theoretical weight of the cake.

Mean and variance of the variable of interest in the case of an effect indicator are:

$$E(\xi) = E(x_3) / \lambda_{3,1}$$
$$Var(\xi) = \left[Var(x_3) - Var(\delta_{3,3})\right] / \lambda_{3,3}^2 \quad (8)$$

Mean and variance of the variable of interest measured by three cause indicators are:

$$E(\eta) = \gamma_{1,1}E(x_1) + \gamma_{1,2}E(x_2)$$
$$Var(\eta) = (\gamma_{1,1})^2\left[Var(x_1) - Var(\delta_{1,1})\right] + (\gamma_{1,2})^2\left[Var(x_2) - Var(\delta_{2,2})\right] + \gamma_{1,1}\gamma_{1,2}2Cov(x_1, x_2) \quad (9)$$

Note that in the case of the effect indicator, we use the subscript "3" because it is a distinct indicator from x_1 and x_2.

Again, we equate both equations $E(\xi) = E(\eta)$ and $Var(\xi) = Var(\eta)$, to analyze if these descriptive statistics depend on the specification of cause or effect indicators.

$$\gamma_{1,1}E(x_1) + \gamma_{1,2}E(x_2) = E(x_3) / \lambda_{3,1}$$
$$(\gamma_{1,1})^2\left[Var(x_1) - Var(\delta_{1,1})\right] + (\gamma_{1,2})^2\left[Var(x_2) - Var(\delta_{2,2})\right] + \gamma_{1,1}\gamma_{1,2}2Cov(x_1, x_2) = \left[Var(x_3) - Var(\delta_{3,1})\right] / \lambda_{3,1}^2 \quad (10)$$

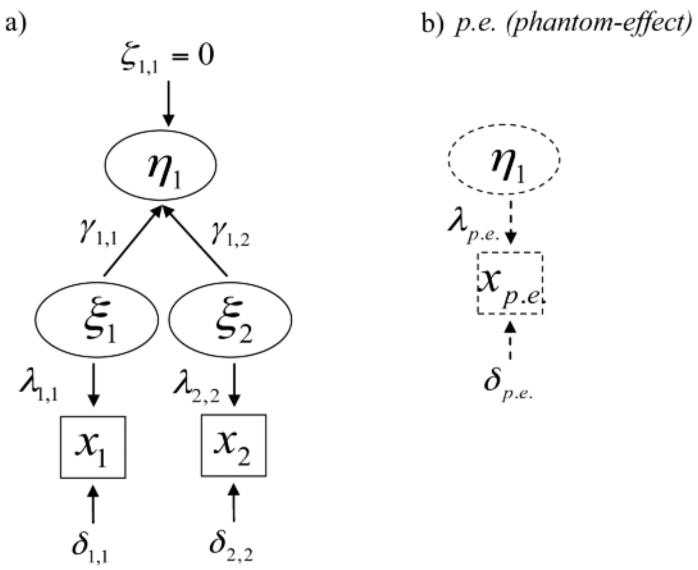
Under the same assumptions as before, quickly it is viewed that $E(\xi) = E(\eta)$, i.e., the expected value of the weight of the theoretical entity is the weight of the cake, or the sum of the weights of its two ingredients. However, again the variance of the variable of interest depends on the measurement specification, but in this case the measurement error of all indicators (effect and cause) is zero. As the variance of the three indicators need not be the same, its error variance can be disparate, even in the case where the indicators were equally reliable. This is the case of the instrument used, the balance, which assumes that relative error carried out in each measure is equal. Consequently, it would be preferable to measure the weight of the cake by using the direct measure of the weight of the cake than the measure of the weight of its ingredients, because measurement error increases as the number of indicators increase.

Formative variables in causal analysis: The use of phantom-effect indicators

Our notion about formative variables has important implications for causal analysis, i.e., when modelling the variable of interest in a causal chain. The theoretical entity can act as a cause or a consequence of other variables in a nomological network. This is also a treated issue in the literature about formative* variables (e.g., Williams, Edwards and Vandenberg, 2003).

When special cases of causal specification occur, operating with formative variables in causal analysis requires the use of what we have called "phantom-effect indicators." A phantom-effect indicator is an algebraic construction derived from the cause indicators, i.e., the weighted composite of cause indicators. This type of indicators is specified as effect indicators of the formative variable, but they do not have the same meaning as the common effect indicators. Variations in the formative variable do not cause variation in the phantom-effect indicator, because the indicators that form the phantom-effect indicator cause variations in the formative variable. These types of indicators are phantom, because they are a numerical trick used to correctly identify some type of causal models. A phantom-effect indicator contains measurement error, as it is the sum of the measurement errors of each cause indicator (Figure 2)

Figure 2
A formative variable measured by cause indicators (a) and by a phantom-effect indicator (b)



Note that the transformation of the formative variable from Figure 2a to Figure 2b is equivalent (11). However, we stress that this does not mean that the formative variable can be measured in a reflective way, because x_3 is not an effect indicator, but a composite of cause indicators, and does not exist separately from cause indicators.

$$\begin{aligned} Var(\delta_{3,3}) &= Var(\delta_{1,1}) + Var(\delta_{2,2}) \\ Var(x_3) &= Var(x_1 + x_2) \\ Var(\eta_1) &= Var(x_3) - Var(\delta_{3,3}) \\ E(\eta_1) &= E(x_3) \end{aligned} \tag{11}$$

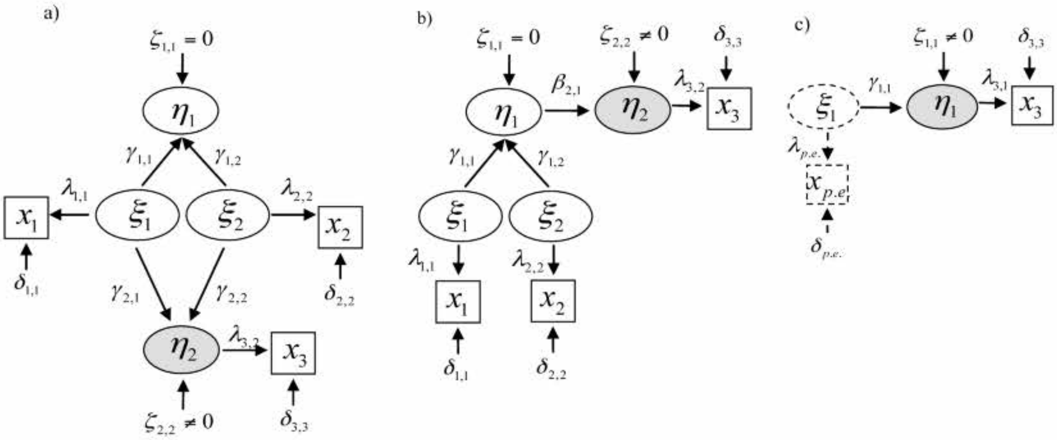
Phantom-effect indicators are necessary to specify some type of causal models. We subsequently explain how to model formative variables in causal analysis.

Formative variable as a cause

This is the case when researchers want to analyze the influence of the variable of interest on other variables. For example, which is the influence of basketball player performance on MVP (most valuable player) awards? Or which is the influence on SES on incidence of low birth weight? Or which is the influence on company performance on CEO turnover?

Our definition of formative variable makes its role easy in this type of analysis. Therefore, the researcher has two options: (1) to consider the cause indicators as causes (Figure 3a) or (2) to consider the algebraic construction as a cause (Figure 3b)

Figure 3
Formative variables as a cause



As Figure 3 shows, the disturbance of the formative variable is zero, according to our previously explained reasoning. We denote η_2 as the variable that acts as a consequence. If cause indicators are considered as causes (in a spurious model), $\gamma_{2,1}$ and $\gamma_{2,2}$ can be estimated using structural equation modelling, assigning a scale to each latent variable and fixing the measurement error of each latent variable (Hayduk, 1996). Therefore, we can know the influence of each cause indicator on η_2. If the formative variable is considered as a cause, we must build an index, as a composite of the two cause indicators. This index has the expected value and the variance showed in equation (4). Using this equation, we may deal with measurement error in the cause indicators. Therefore, by fixing the measurement error in cause indicators, we can compute the measurement error in the index, and thus we can estimate $\beta_{2,1}$.

Using the model in Figure 3a, we can determine the relative contribution of each cause indicator to the variance of η_2, so it provides a more detailed view of the influence of the formative variable on η_2. As the formative variable depends on its definition, i.e., the choice of the cause indicators, rival measurement models could be tested to compare several definitions of the formative variable. For example, researchers can compare the numerous basketball player performance measurement models that have been proposed, to analyze which model better explains η_2, and to study which are the cause indicators (e.g., points, rebounds, assists) that are the most important predictors of η_2.

Model in Figure 3b can be transformed to the model in Figure 3c, using a phantom-effect indicator. Both models yield the same estimated effect of the formative variable on the other variable, so $\beta_{2,1}$ in Figure 3b would be equal to $\gamma_{1,1}$ in Figure 3c. In addition, both models would explain the same amount of variance of the dependent variables, i.e. $\zeta_{2,2}$ in Figure 3b would be equal to $\zeta_{1,1}$ in Figure 3c.

Formative variable as a consequence

This is a case when researchers want to analyze the influence of other variables on the variable of interest. For example, what is the influence of ISO certification on company performance? Again, we can consider two options: (1) the cause indicators as consequences (Figure 4a) or (2) the algebraic construction as a consequence (Figure 4b).

Conclusion

The aim of this paper was to propose a new understanding of formative measurement models: *in a formative measurement model in which a construct is defined in terms of its measurements; the variable of interest is a composite of indicators, an algebraic construction that cannot be measured in a reflective way*. In addition, formative variables have no disturbance, i.e., there is no error term, because the definition of these variables exclusively depends on its indicators, which fully determine the meaning of the formative variable. If other indicators are specified as cause indicators of this formative variable, then the meaning of the formative variable would be different; so the new formative variable would be a different variable from the previous one.

However, our understanding of formative variables allows counting for measurement error in cause indicators, because the spurious model re-specifies cause indicators to latent variables measured in a reflective way. These latent variables form the formative variable, which is also a latent variable that does not exist separately from its indicators, but has a different meaning from its indicators.

This definition of formative variables breaks down with the current understanding of formative* variables, and it guarantees that they are not empirically equivalent to variables measured by using reflective indicators. In addition, it avoids the problem of interpretational confounding.

We proposed that variables can be classified into three categories: cake, love, and player performance. Formative variables fall into the player performance category, i.e, algebraic constructions derived from other variables, which cannot be measured in a reflective way. They are defined in terms of their indicators but have their own meaning. Cake and love variables can be measured with either cause or effect indicators. Algebraic constructions can be formed, creating an index, but this index does not have the same meaning as the index corresponding to a formative variable, because the latter cannot be empirically equivalent to a reflective measure of the same variable. In fact, formative variables cannot be reflectively measured.

However, formative variables can be re-specified using phantom-effect indicators. A phantom-effect indicator is an algebraic construction derived from the cause indicators, and it is specified as an effect indicator of the formative variable, but it does not have the same meaning as the common effect indicators. These types of indicators are phantom, because they are a numerical trick used to correctly identify some types of causal models.

We expect that these new thoughts regarding formative variables would assist researchers in implementing adequate measures of what they want to measure. We recommend thinking about the three categories we proposed to select cause or effect indicators. In addition, we encourage researchers to critically think regarding the use of formative variables in causal models, using phantom-effect indicators if necessary.

This research does not fully cover all the subjects that are a matter of discussion in the specialized literature about measurement. In fact, we have not considered the very recent proposal of Hayduk, Pazderka-Robinson, Cummings, Boadu, Verbeek, and Peerks (2007) regarding reactive indicators, a class of indicators that act as both the cause and effect of an underlying latent variable, because we believe that this requires a more deep philosophical discussion, which is beyond the objectives of our research. However, we hope that our understanding of formative variables will improve the current thinking about this theme, and provide a general framework to operate in empirical research.

References

APGAR, V. (1953). "*A proposal for a new method of evaluation of the newborn infant*". Current Researches in Anaesthesia & Analgesia, 32 (4), 260–267

BAGOZZI, R. P. (2007). "*On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox*" Psychological Methods, 12 (2), 229-237.

BOLLEN, K. A. (2002). "*Latent variables in psychology and the social sciences*". Annual Review of Psychology, 53, 605-34

BOLLEN, K. A. (2007). "*Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox*". Psychological Methods, 12 (2), 219-228.

BORSBOOM, D., MELLENBERGH, G. J., & VAN HEERDEN, J. (2003). "*The theoretical status of latent variables*". Psychological Review, 110 (2), 203-219.

BURT, R. S. (1976). "*Interpretational confounding of unobserved variables in structural equation models*". Sociological Methods Research, 5, 3–52.

CASEY, B. M., MCINTIRE, D. D., & LEVENO, K. J. (2001). "*The continuing value of the Apgar score for the assessment of newborn infants*". New England Journal of Medicine, 344 (7), 467-471.

COLDITZ, G. A., ATWOOD, K. A., EMMONS, K., MONSON, R. R., WILLETT, W. C., TRICHOPOULOS, D., & HUNTER, D. J. (2000). "*Harvard report on cancer prevention. Volume 4: Harvard Cancer Risk Index*". Cancer Causes and Control, 11, 477-488

COLTMAN, T., DEVINNEY, T. M., MIDGLEY, D. F., & VENAIK, S. (2008). "*Formative or reflective scales: Two applications of erroneous measurement*". Journal of Business Research, 61 (12), 1250-1262.

DIAMANTOPOULOS, A. (2006). "*The error term in formative measurement models: interpretation and modeling implications*". Journal of Modeling in Management, 1 (1), 7–17

DIAMANTOPOULOS, A., & WINKLHOFER, H. (2001). "*Index construction with formative indicators: An alternative to scale development*". Journal of Marketing Research, 38 (2), 269-277.

EDWARDS, J. R., & BAGOZZI, R. P. (2000). "*On the nature and direction of relationships between constructs and measures*". Psychological Methods, 5 (2), 155-174.

HAYDUK, L. A. (1996). LISREL Issues, Debates and Strategies. Baltimore, MD.

HAYDUK, L. A., PAZDERKA-ROBINSON, H., CUMMINGS, G. C., BOADU, K., VERBEEK, E, L., & PEERKS, T. A. (2007). "*The weird world, and equally weird measurement models: Reactive indicators and the validity revolution*". Structural Equation Modeling, 14 (2), 280-310.

HOLLINGER, J. (2005). Pro Basketball Forecast: 2005-2006. Potomac Books Inc

HOWELL, R. D., BREIVIK, E., & WILCOX, J. B. (2007). "*Reconsidering formative measurement*". Psychological Methods, 12 (2), 205-218.

JARVIS, C. B., MACKENZIE, S. B., & PODSAKOFF, P. M. (2003). "*A critical review of construct indicators and measurement model misspecification in marketing and consumer research*". Journal of Consumer Research, 30 (2), 199-218.

MARKUS, K. A. (1998). "*Science, measurement, and validity: Is completion of Samuel Messick's synthesis possible?*". Social Indicators Research, 45, 7-34.

MARKUS, K. A. (2004). "*Varieties of causal modeling: How optimal research design varies by explanatory strategy*". In K. van Montfort, J. Oud and A. Satorra (eds) Recent Developments on Structural Equation Models: Theory and Application: Dordrecht: Kluwer Academic Publishers.

OLIVER, D. (2004). Basketball on paper. Rules and tools for performance analysis. Washington, D.C.: Brassey's, Inc.

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.

Punset, E. (2007). "*El viaje al amor*". Destino, Barcelona.

Rossiter, J. R. (2002). "*The C-OAR-SE procedure for scale development in marketing*". International Journal of Research in Marketing, 19 (4), 305–335.

Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). "*Questions about formative measurement*". Journal of Business Research, 61, 1219-1228.

Williams, L. J., Edwards, J. R., & Vandenberg, R. J. (2003). "*Recent advances in causal modeling methods for organizational and management research*". Journal of Management, 29 (6), 903-936.