Measurement residual theory, analytic composites, and the assessment of formative latent variables

Ned Kock

Full reference:
Kock, N. (2025). Measurement residual theory, analytic composites and the assessment of formative latent variables. *Journal of Systems and Information Technology*, 27(2), 310-340.

Abstract

Purpose
Formative measurement, where indicators are frequently seen as causing their corresponding latent variables, is extensively used in information systems research; and in such a way as to attract methodological criticism to the entire field. We attempt to ameliorate this situation.
Design/methodology/approach
Anchored on a new measurement residual theory, we argue that a latent variable always exists before the corresponding indicators when data is collected via questionnaires, whether reflective or formative measurement is used. Consequently, we posit that the direction of causality going from indicators to latent variables normally associated with formative measurement is misguided.
Findings
We develop a theory-driven set of recommendations for the assessment of formative measurement quality, addressing the following elements: factor reliability, indicator redundancy, significance of indicator weights, indicator effect sizes, Simpsons' paradox instances associated with indicators, model-wide factor redundancy, and use of analytic composites.
Research limitations/implications
The new theory and related recommendations are illustrated based on an empirical study of 290 geographically distributed product innovation teams that used various electronic communication media to conduct their work.
Originality/value
The data is analyzed with the software WarpPLS, a widely used structural equation modeling tool that allows for formative measurement assessment and analytic composite utilization, in ways that are fully compatible with the theory-driven set of recommendations presented in this paper.

Keywords: Measurement Residual Theory; Common Factor Model; Measurement Error Theory; Structural Equation Modeling; Latent Variable; Analytic Composite

1. Introduction

Questionnaires are developed by researchers to indirectly measure unobservable variables, frequently called latent variables, which are mental ideas – e.g., job satisfaction and e-communication media use. These latent variables, also referred to as constructs and latent constructs, are assumed to be quantified through factors. Question-statements typically answered on Likert-type scales produce indirect measures of the mental ideas when answered. These indirect measures are known as indicators, and are believed to individually measure latent variables with some degree of imprecision.

There are two main types of latent variable measurement: reflective and formative (Bollen, 2011; Cenfetelli & Bassellier, 2009; Cheah et al., 2019; Coltman et al., 2008; Fleuren et al., 2018; Hanafiah, 2020; Kono et al., 2021; Mikulić & Ryan, 2018; Theodosiou et al., 2019; Peštović et al., 2021). Reflective measurement relies on redundant question-statements, while formative measurement involves the use of non-redundant question-statements. Let us say that a researcher wants to measure the use of e-communication media by individuals in organizations. Question-statements of the type "I use e-communication media" and "using e-communication media is important to me" would be used in reflective measurement. Question-statements of the type "I use email" and "I use video conferencing" would be used in formative measurement.

As we can see above, in formative measurement the question-statements are non-redundant, but they are generated based on one mental construct – use of e-communication media. In the example above, the question-statements are not redundant because heavy email users may not be heavy video conferencing users, and vice-versa. Without the mental, or latent, construct (i.e., e-communication media use), it would not be possible to generate meaningful question-statements. Meaningless question-statements, like “I use or do not use whatever”, would be of little use for measurement.

In spite of this, there is a general belief among scholars that, with formative measurement, indicators cause their corresponding latent variables. This idea, which is discussed in more detail in this paper, is clearly illogical, because it requires the existence of the indicators prior to the latent variables that they are designed to measure – and those latent variables are essentially the mental ideas needed to generate meaningful question-statements. The solution to this conundrum requires, in our view, a reconceptualization of formative measurement, which we propose in this paper, along with a set of guidelines on how to assess the quality of such measurement.

2. Research background and goal

In formative measurement the indicators are often thought of as causes of the latent variables, and by extension of the associated factors (Hardin, 2017; Ho et al., 2022; Giovanis et al., 2018; Hardin et al., 2011; Hsu et al., 2018). But how indicators based on question-statements developed to indirectly measure latent variables can possibly cause those variables? After all, the mental ideas associated with the latent variables must exist before the question-statements are developed.

In part due to complications associated with this paradoxical situation, formative measurement has been heavily criticized in the past. For example, Hardin & Marcoulides (2011) call for suspending the use of formative measurement. Edwards (2011) goes one step further, and calls for a complete abandonment of formative measurement. Howell et al. (2013), too, put forth an equally critical perspective (see, also: Howell et al., 2007a; 2007b). Our view, explained in this paper, is that formative measurement can indeed be used, as long as it is properly conceptualized and certain precautions are taken.

Formative measurement is used in information systems research in such an extensive way as to attract criticism to the entire field. Hardin & Marcoulides (2011, p. 753), in their call for suspending the use of formative measurement, noted that a “flurry of articles on formative measurement, particularly in the information systems literature, appears to be symptomatic of a much larger problem. Despite significant objections by methodological experts, these articles continue to deliver a predominately pro formative measurement message to researchers who rapidly incorporate these recommendations into their research”.

The heavy use of formative measurement in the field of information systems may have been motivated by the popularity of partial least squares algorithms (Diamantopoulos, 2011; Kim et al., 2010; Kock, 2019; 2023; Petter, 2018; Petter et al., 2007), particularly the algorithm variation known as Mode B, also called the “formative mode” (Lohmöller, 1989). Even though it is widely used, the problematic nature of Mode B, leading to interpretational difficulties such as Simpson’s paradox instances (Kock, 2014; 2021b; Pearl, 2009), has been clearly demonstrated in the past; with compelling demonstrations having been provided by information systems researchers (Aguirre-Urreta & Marakas, 2013; Kock & Mayfield, 2015).

The above scenario creates difficulties for the use and assessment of formative measurement, and consequently a significant gap in the methodological literature. Hence, our goal in this paper is to fill this gap though a reconceptualization of formative measurement, which requires the development of a new theory of indirect measurement of latent variables with error via indicators, and the development of a set of guidelines on how to assess the quality of such measurement that is directly derived from this new theoretical framework.

We argue in this paper, anchored on a new measurement residual theory, referenced through the acronym MRT, that the indicators-to-factor causal links direction normally associated with formative measurement is misguided. Moreover, we propose a set of recommendations for the assessment of formative measurement quality, addressing the following elements: factor reliability (Canatay et al., 2022; Kock, 2023), indicator redundancy (Kock, 2014; Kock & Moqbel, 2021; Petter et al., 2007), significance of indicator weights (Amora, 2023; Kock, 2014; Petter et al., 2007), indicator effect sizes (Amora, 2023; Kock, 2014), Simpsons' paradox instances associated with indicators (Amora, 2023; Kock, 2014; 2021b; Pearl, 2009), model-wide factor redundancy (Bayonne et al., 2020; Kock & Lynn, 2012), and use of analytic composites (Bag et al., 2022; Kock, 2021a). Some of these recommendations are novel and add important elements to existing formative measurement assessment criteria; e.g., when Simpsons' paradox occurs at the indicator level, the indicator in question makes a *negative* individual contribution to the variance explained in the factor (Kock, 2022; Kock & Gaskins, 2016; Pearl, 2009).

For simplicity, and without any impact on the generality of our discussion, all of the variables we use in our argumentation are assumed to be standardized. That is, unless stated otherwise, the variables are assumed to be scaled to have a mean of zero and standard deviation of one. Moreover, the parameters to which we refer (e.g., regressions weights) are also assumed to be standardized parameters, unless stated otherwise. Finally, to simplify our discussion at some points, and avoid long-winded sentences, we use certain terms that refer to closely related entities interchangeably. Examples are the terms latent variable and factor, and the terms question-statement and indicator.

3. Measurement residual theory

In this section, and component sub-sections, we lay out a new “summative” conceptualization of measurement residuals and related theoretical elements. This new theory, to which we refer by employing the acronym MRT, standing for measurement residual theory, incorporates key elements from two well established theoretical frameworks. These frameworks are classic measurement error theory (Nunnaly, 1978; Nunnally & Bernstein, 1994) and the common factor model (Kline, 2010; MacCallum & Tucker, 1991). MRT refers primarily to measurement error arising from the use of questionnaires, leading to measurement residuals that differentiate factors from composites (Kock, 2015b; 2019).

We believe that MRT makes a notable contribution to the literature by clarifying key issues that underlie structural equation modeling with both formative and reflective latent variables, through the novel integration of classic measurement error theory and the common factor model. We also believe that MRT makes another important contribution to the literature by providing an integrated theory that serves as a basis on which one can build recommendations for the assessment of formative measurement quality, as we do. We hope it will be clear in this paper that, without MRT, our recommendations would not flow in a very logical and theory-driven fashion. In this sense, MRT provides a guiding framework that we hope will be useful in future methodological discussion by other researchers.

3.1. A factor always causes its indicators

Let us say that a researcher wants to measure job satisfaction, so that a study can be conducted to investigate whether one’s job satisfaction influences job performance at an organization. In this example, job satisfaction is an idea that exists in the mind of the researcher *before* any question or question-statement is devised to measure the construct. The researcher may devise the following question-statements, to be answered on a Likert-type scale with 7 points (e.g., 1 = very strongly disagree, 2 = strongly disagree … 7 = very strong agree) to measure job satisfaction: "I like my job", and "My job is great". The sets of scores obtained based on these two question-statements across multiple individuals are referred to as *indicators* of the job satisfaction construct.

In the example above, a factor is a set of scores that is assumed to measure the mental idea of job satisfaction, a construct that is not directly observable. The indicators are not available to the researcher prior to an empirical study, where a questionnaire is administered to study participants. The question-statements above are expected to measure one *single dimension* of job satisfaction, which is why the question-statements appear to be redundant (i.e., ask the same "thing" using different words). As such, the corresponding indicators are said to be associated with the factor in a *reflective* way (Amora, 2023; Kock, 2014; Mohammad Salameh et al., 2018). This is a simplified example for illustration purposes; usually more than two question-statements would be used.

The researcher may instead use the following question-statements to measure the job satisfaction construct: "I like my boss", and "I like my office". These question-statements are different in that they can reasonably be expected to measure two different dimensions of job satisfaction, namely satisfaction with one's boss and with one's office. As such, they are said to measure the job satisfaction construct, or to be associated with the corresponding factor, in a *formative* way (Ambalov, 2021; Kock, 2014; Mamakou et al., 2024).

In either case, reflective or formative, the job satisfaction idea must exist in the mind of the researcher before the question-statements are devised (see Figure 1). Therefore, in both cases the factor, which is assumed to measure the corresponding construct, comes first from a temporal perspective. In other words, one could say that the factor *always* causes the indicators, whether measurement is reflective or formative, since the factor is an unobserved measure of the latent construct.

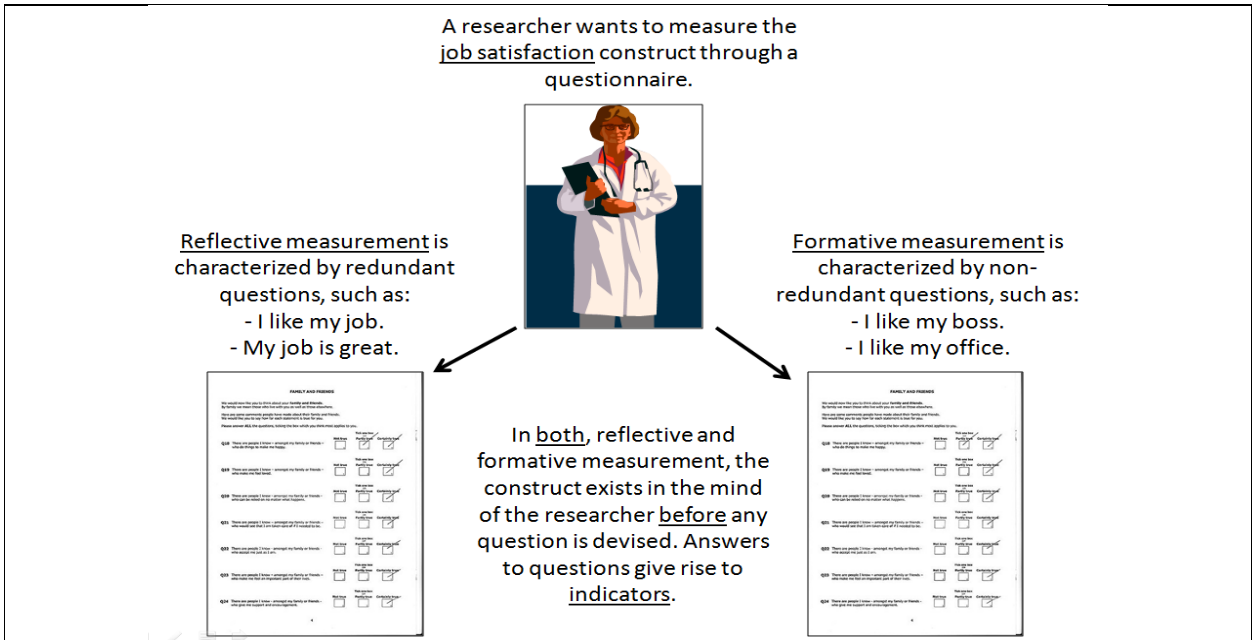

Figure 1. Latent constructs exist before their indicators

Another construct, which was used earlier in this paper and would be particularly relevant in the field of information systems, is the use of e-communication media by individuals in organizations; incidentally, this construct has been shown to be related to both job satisfaction and job performance (see, e.g., Kock & Moqbel, 2021). Question-statements of the type "I use e-communication media" and "using e-communication media is important to me" would be used in reflective measurement. Question-statements of the type "I use email" and "I use video conferencing" would be used in formative measurement.

3.2. The measurement residual of a factor

Unlike classic measurement error theory and the common factor model, MRT conceptualizes the measurement residual associated with a factor as akin to a “unique indicator” that accounts for the variance in the factor that is not explained by the indicators associated with the factor. This measurement residual should not be confused with the indicator errors, which are key elements of the common factor model. Figure 2 illustrates these several conceptual entities and corresponding relationships hypothesized by MRT.

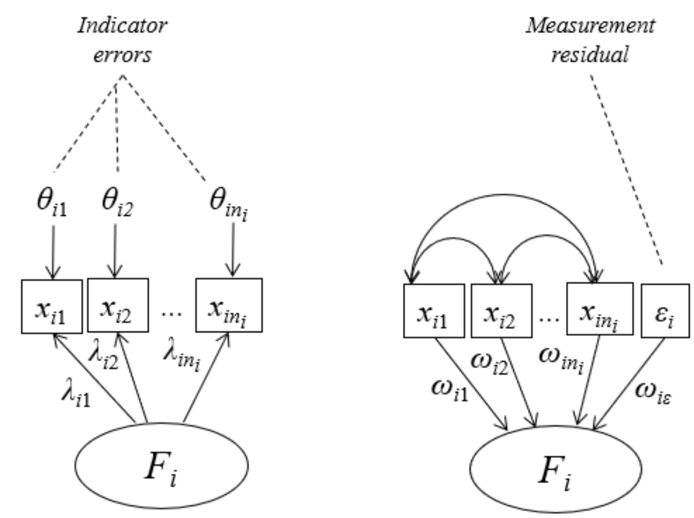

Figure 2. Indicator errors and measurement residual of a factor

Notes: arrows used in a schematic way, to indicate whether indicators are regressed on their factor (left), or vice-versa (right); true causality is always from factor to indicators when questionnaires are used.

The arrows are used here in a schematic way, to indicate whether indicators are regressed on their factor, or vice-versa, to obtain loadings and weights respectively. As noted earlier, true causality always goes from factor to indicators when questionnaires are used for data collection and subsequent analysis.

In this figure $x_{i1}, x_{i2} \ldots x_{in_i}$ are the n_i indicators of a factor F_i; and $\lambda_{i1}, \lambda_{i2} \ldots \lambda_{in_i}$ the corresponding standardized loadings, which are the correlations among the factor and each of the indicators. The assumed individual imprecision of each of the indicators implies individual loadings lower than 1, and the resulting existence of indicator errors θ_{i1}, θ_{i2} ... θ_{in_i}. In MRT the partial regression weights ω_{i1}, ω_{i2} … ω_{in_i} are hypothesized to arise when the factor is regressed on its indicators, in a least squares sense. The factor's measurement residual ε_i is hypothesized in MRT to arise as an element that accounts for the variation in the factor that is not accounted for by the indicators.

All of the entities above are assumed in MRT to be stored in column vectors, and to exist at the population level. In this sense, the entities above are assumed to be the *true* entities that satisfy the stated assumptions underlying MRT. This applies to both, left and right, parts of the figure. For example, the weights ω_{i1}, ω_{i2} … ω_{in_i} in the figure are the *true weights*. The population is assumed to have a finite size N_p, which can be quite large, e.g., the entire human population; or small, e.g., 50 individuals who work at a small organization with unique characteristics. When a sample that is smaller than the population is randomly drawn from it, it will have a size N. Samples generated by empirical studies are assumed to replicate this random sampling process.

3.3. The true composite associated with a factor

Composites, unlike factors, are exact linear combinations of indicators that do not incorporate the variation found in the corresponding factors' measurement residuals. As such, composites can be seen as approximations of factors, whether measurement is reflective or formative. There are many ways in which composites used to approximate factors can be created; e.g., by using identical weights or weights generated by classic partial least squares algorithms (Adelman & Lohmoller, 1994; Kline, 2010; Kock, 2019; 2023; Lohmöller, 1989; McIntosh et al., 2014).

In MRT, however, only one unique composite is hypothesized to arise from the weighted aggregation of the indicators using the true weights $\omega_{i1}, \omega_{i2} \dots \omega_{in_i}$. This is referred to as the *true composite* C_i. This entity's relationship with its corresponding factor F_i is indicated in Eq. 1, where x_i is a matrix storing the indicators associated with factor F_i, ω_i is a column vector storing indicator weights, and ω_{iC} is the composite weight.

$$\begin{gathered} F_i = \sum_{j=1}^{n_i} x_{ij}\omega_{ij} + \varepsilon_i\omega_{i\varepsilon} \rightarrow \\ F_i = x_i\omega_i + \varepsilon_i\omega_{i\varepsilon} \rightarrow \\ F_i = C_i\omega_{iC} + \varepsilon_i\omega_{i\varepsilon}. \end{gathered} \tag{Eq. 1}$$

In Appendix A we demonstrate that the true composite weights must satisfy Eq. 2, where: $\Sigma_{x_i x_i}$ is the covariance matrix of the indicators; $\Sigma_{x_i \theta_i}$ is the matrix of covariances among indicators and their errors; the function $diag(\cdot)$ returns a matrix with only its diagonal elements different from zero; the superscript $'$ denotes the transpose operation; the superscript -1 denotes the classic matrix inversion; and the superscript $+$ denotes the Moore–Penrose pseudoinverse transformation. We should point out that the weights stored in ω_i are not, to our knowledge, estimated by any of the widely used algorithms currently employed to estimate composites; this includes classic partial least squares algorithms (see, e.g., Lohmöller, 1989).

$$\omega_i = \Sigma_{x_i x_i}{}^{-1}\left(\Sigma_{x_i x_i} - diag\left(\Sigma_{x_i \theta_i}\right)\right)\lambda_i{}'^{+}. \tag{Eq. 2}$$

Since each measurement residual ε_i is uncorrelated with the indicators stored in x_i, we can follow a simple stochastic set of steps to produce the estimates of the true weights ($\widehat{\omega}_i$) using Eq. 2. With these weight estimates and the indicators stored in x_i we can then directly obtain estimates of the true composites. The stochastic set of steps is as follows. Step 1: Initialize $\hat{\varepsilon}_i$

with random values; a stochastic step. Step 2: Produce estimates of the loadings ($\hat{\lambda}_i$) and reliability ($\hat{\rho}_i$) using previously validated approaches such maximum likelihood confirmatory factor analysis (Kline, 2010; Mueller, 1996), or the consistent partial least squares technique (Dijkstra & Schermelleh-Engel, 2014). Step 3: Calculate $\widehat{\omega}_{iC}$ and $\widehat{\omega}_{i\varepsilon}$ based on the reliability estimate $\hat{\rho}_i$. Step 4: Initialize $\widehat{\omega}_i$ with unit values. Step 5: Perform iterations using Eq. 2 until the weights in $\widehat{\omega}_i$ change by less than a small fraction. Kock (2015b) provides an illustration and validation of a variation of this set of steps.

3.4. The true reliability associated with a factor and correlation attenuation

Since the measurement residual is assumed to be uncorrelated with the indicators of a factor in MRT, it is by definition uncorrelated with the true composite. Therefore, we can conclude based on classic measurement error theory that the true reliability associated with each factor F_i is the amount of variance explained by the true composite. It should be noted that this is the same as the amount of variance explained by the indicators. In MRT we refer to this true reliability as ρ_i. It thus follows that the weights associated with the true composite and measurement residual can be obtained directly from the true reliability:

$$\omega_{iC} = \sqrt{\rho_i},$$
$$\omega_{i\varepsilon} = \sqrt{1 - \rho_i}.$$

Classic measurement error theory states that, in the presence of measurement error, the correlation between any pair of composites $\Sigma_{C_iC_j}$ has a lower absolute magnitude than the correlation between the corresponding factors $\Sigma_{F_iF_j}$ (Nunnaly, 1978; Nunnally & Bernstein, 1994). The correlation between any pair of composites equals the correlation between the corresponding factors multiplied by the geometric mean of the factors' true reliabilities:

$$\Sigma_{C_i C_j} = \Sigma_{F_i F_j} \sqrt{\rho_i \rho_j}. \quad \text{(Eq. 3)}$$

It should be noted that Eq. 3 only holds if the true values of the correlations between factors, the correlations between the true composites, and the true reliabilities are used. Fairly accurate estimates of the true reliabilities can be obtained via maximum likelihood confirmatory factor analysis (Kline, 2010; Mueller, 1996), and the consistent partial least squares technique (Dijkstra & Schermelleh-Engel, 2014). With those, and the equations discussed in MRT for the estimation of the true composite scores, one can obtain via Eq. 3 the true correlations among factors. Many other entities and parameters can be subsequently estimated; including the factors themselves, via the Thurstone and Bartlett methods (DiStefano et al., 2009; Bartlett, 1937; Thurstone, 1935), as well as the new variation sharing method (Kock, 2017; Kock & Sexton, 2017). It should be noted that the factor estimates generated via the Thurstone and Bartlett methods have been found recently to be rather imprecise (Kock, 2019).

In MRT a measurement residual always exists in connection with a factor, whether measurement is formative or reflective. This follows the conventional view of formative measurement, of which seminal discussions are provided by Bollen & Lennox (1991) and Diamantopoulos (2011). In it, formative measurement assumes the existence of a residual, which is not present in composites. In MRT factors are not modeled as composites; they are weighted aggregations of their indicators and measurement residual. In this context, the true reliability is the percentage of the variance in the factor that is explained by the indicators. This applies to both reflective and formative measurement, and it goes generally against the composite-based orientation underlying classic partial least squares techniques (Lohmöller, 1989). In MRT the true composite and factor are closely connected but distinct entities.

4. Analytic composites

Many indices are used in business and societal contexts that are aggregations of other measures, and that do not explicitly incorporate measurement error. From a mathematical statistics standpoint, these are composites. Examples are the Standard & Poor's 500 and the Dow Jones Industrial Average, which are stock market indices; and the Gini coefficient, a country wealth distribution index. To avoid confusion with the term index as it is used in structural equation modeling, normally referring to a measure of fit (Kline, 2010; Mueller, 1996), we will refer to these as *analytic composites*.

One common characteristic of analytic composites is that they are *designed* to serve a purpose. For example, the Standard & Poor's 500 has been designed to provide a key piece of information to the public, namely whether the U.S. stock market as a whole is "going up or down" in terms of company valuations. The weights of the indicators used in the Standard & Poor's 500 are based on the market capitalizations of 500 large companies listed on two major stock exchanges, the NYSE and NASDAQ. Unlike the weights associated with true composites and factors in our formulation of MRT, which are estimated through mathematical equations based on the indicators, the weights in analytic composites such as the Standard & Poor's 500 are set by the designers of the analytic composites.

A researcher can use indicators taken from a questionnaire to build an analytic composite with pre-specified indicator weights. For example, let us assume that an analytic composite A_1 is to be built with indicators x_{11}, x_{12} and x_{13} from a questionnaire. The weights to be used are respectively: .665, .333, and .166. Note that the first weight is approximately twice the second, which is twice the third. The composite would thus be calculated as indicated in Eq. 4. Appendix B shows how an analytic composite can be built in practice, in the context of a structural equation modeling analysis.

$$A_1 = x_{11}(.665) + x_{12}(.333) + x_{13}(.166). \quad \text{(Eq. 4)}$$

For this type of design to be useful in the context of a research study, the researcher must have good reasons to build the analytic composite in a particular way; i.e., using pre-specified indicators and corresponding weights. Within the scope of our discussion on formative measurement, analytic composites can be useful as aggregators of indicators that do not pass quality assessment criteria; e.g., indicators whose factor weights are individually too small to be considered statistically different from zero. One such analytic composite is indicated as A_i in Eq. 5.

$$\begin{gathered} F_i = \sum_{j=1}^{10} x_{ij}\omega_{ij} + \varepsilon_i\omega_{i\varepsilon} \rightarrow \\ F_i = \sum_{j=1}^{5} x_{ij}\omega_{ij} + \sum_{j=6}^{10} x_{ij}\omega_{ij} + \varepsilon_i\omega_{i\varepsilon} \rightarrow \\ F_i = \sum_{j=1}^{5} x_{ij}\dot{\omega}_{ij} + A_i\omega_{iA} + \varepsilon_i\dot{\omega}_{i\varepsilon}. \end{gathered} \quad \text{(Eq. 5)}$$

Here we assume that a factor F_i refers to formative measurement and aggregates 10 indicators, of which only the first 5 have weights that are individually strong enough to be statistically significant – or assumed to be nonzero weights from a statistical standpoint. The other 5 indicators have weights that are individually too weak, but that when aggregated into the composite A_i based on their original weights lead to a composite weight ω_{iA} that is statistically significantly. This happens while the other 5 indicators retain significant weights, even though those weights change (from ω_{ij} to $\dot{\omega}_{ij}$). This analytic composite could be seen as analogous to an instrumental variable (Arellano & Bover, 1995; Kline, 2010) that implements an instance of second-order measurement. Nevertheless, it should be clear that it does not interfere with the incorporation of measurement error into the factor.

As it will be seen later, such an auxiliary analytic composite can help us realize formative measurement that passes a comprehensive set of measurement quality criteria, proposed by us consistently with MRT. The main underlying reason why this type of analytic composite would be used in formative measurement is that it may incorporate variation that can be useful in that measurement, and that would otherwise be lost in a structural equation modeling analysis. Without that variation, a formative factor may end up with too much of its variance being explained by its corresponding measurement residual, which would likely lead to a downward bias in path coefficients for links causally connecting the factor with other factors (Kline, 2010; Kock, 2019).

Pseudo-reliability estimates can be produced for an analytic composite through the classic equations defining the Cronbach's alpha (see, e.g.: Nunnaly,1978; Nunnally & Bernstein, 1994) and the composite reliability (see, e.g.: Dillon & Goldstein, 1984; Peterson & Yeolib, 2013). These pseudo-reliability estimates should not be confused with the true factor reliability, which is the amount of variance explained by the indicators associated with the factor. Strictly speaking, all composites have a reliability of 1, because 100 percent of their variance is explained by their component indicators.

5. Formative or reflective? Semantic assessment and the loadings rule

While in formative measurement the researcher normally expects different indicators to measure diverse dimensions of the same latent variable, this may not be the case if the questionnaire respondents do not construe the question-statements in the same way as the researcher who designed the question-statements did. For example, let us assume that a latent variable associated with the use of project management techniques by teams is measured through the following question-statements: “There were adequate mechanisms to track the project's progress”, and “There were adequate mechanisms to track the project's costs”. It is possible that a researcher would see these questions as measuring different dimensions of a latent variable, whereas the questionnaire respondents could subconsciously view them as redundantly measuring the same construct, or vice-versa. Another possibility is that the mechanisms available for tracking a project’s progress and costs are always used by teams together, even as the degrees to which they are used vary across teams, leading the indicators to be redundant.

Given the above possibilities, we argue that the decision as to whether to consider a measurement approach as formative or reflective, for measurement quality assessment purposes, can be made only after data is collected and analyzed, with the results being compatible with theoretical assumptions made by the researcher at the questionnaire design stage. This latter requirement, which refers to the questionnaire design stage, often can be tested via commonsense, and can be illustrated through an example. An indicator associated with the question-statement “I like my spouse” cannot, under most circumstances, be theoretically associated with the mental idea of the market success of a new product developed by a team at an organization. Below we propose a more structured approach to address this type of problem, through a semantic assessment check. We follow that with the proposal of a general decision rule building on indicator loadings.

5.1. Semantic assessment

Formative and reflective measurement are often confused (Bollen & Lennox, 1991; Hsu et al., 2018). A semantic assessment check, focused on the *meaning* of the question-statements used for data collection (Bollen, 2011; Diamantopoulos, 2011), is of critical importance to avoid a situation that appears to be all too common (Cenfetelli & Bassellier, 2009; Petter et al., 2007), where poorly designed reflective measurement is confused with formative measurement. This goes beyond the commonsense validations mentioned above. For this, the unit of analysis for data collection regarding each latent construct must be clearly defined (Kock & Lynn, 2012). Indicators measure latent constructs with respect to a given unit of analysis, which we view as providing an important semantic "anchor" for the latent construct to be quantified. Let us consider the indicator associated with the question-statement "There were adequate mechanisms to track the project's progress". The unit of analysis here is the project, because what is being measured is an attribute of a given project – namely the degree to which adequate mechanisms were being used to track its progress. Note that even though the data may be collected at the individual level (e.g., from project team leaders), the unit of analysis in this example is still the project carried out by the team.

Once the unit of analysis is clearly established for a particular latent construct, a semantic assessment can be conducted by ensuring that all question-statements associated with the same latent construct refer to the same unit of analysis. In this sense, the following question-statements would pass this semantic assessment check, because they both refer to the same unit of analysis (i.e., the project): "There were adequate mechanisms to track the project's progress" and "There were adequate mechanisms to track the project's costs". However, this question-statement would *fail* the semantic assessment check: "There were adequate mechanisms to track the success of the firm where the project was conducted”. The reason is that this question-statement refers to a different unit of analysis, namely the firm where the project was carried out. Indicators associated with question-statements that fail this semantic assessment check should be either removed from the analysis or re-assigned to another latent construct (Kock & Lynn, 2012), if they pass the check with respect to that other construct.

5.2. The loadings rule

Because we anticipate that, in formative measurement, indicators will measure different dimensions of the same latent variable, and thus will not be redundant among themselves, we would expect formative latent variables to fail to conform to the classic convergent validity expectation that loadings be equal to or greater than .5 (Kline, 2010; Kock & Lynn, 2012). This expectation assumes that all the indicators associated with a given latent construct pass a semantic assessment check, as discussed above. Those indicators that fail the check should first be removed from the analysis or re-assigned to another latent construct. After this, loadings should be re-calculated.

After the above is conducted, and consistently with the classic convergent validity expectation discussed above, it is our view that latent variables for which all indicators have loadings equal to or greater than .5 should be treated as reflective, and subject to classic validity and reliability measurement assessment criteria applied to reflective latent variables (Kline, 2010; Kock & Lynn, 2012). If that is not the case, then they should be treated as formative and subject to the recommendations proposed below. We refer to this as the *loadings rule* for the decision to employ formative or reflective treatment of latent variables, with respect to the assessment of their measurement quality.

At this point, an expert reader may correctly argue that the loadings can also be lower than .5 for reasons such as response error (e.g., due to cognitive fatigue or translation errors) or owing to bad sampling, which are reasons that would be unrelated to formative measurement. We believe that, if this is the case, the problems will be uncovered by the combined use of the guidelines below. For example, if a loading is lower than .5 due to response error, it is unlikely that the following three criteria will be simultaneously met: (a) the corresponding indicator will have a full collinearity variance inflation factor of 3.3 or lower; (b) the corresponding indicator weight will be significantly different from zero; and (c) the corresponding indicator effect size will be .02 or greater. These three criteria refer to three of the guidelines discussed in the section that follows.

6. Formative measurement and the assessment of its quality

In the following sub-sections, we provide a set of recommendations for formative measurement and the assessment of the quality of that form of measurement. They refer to the use of causal arrows pointing in or out of factors, factor reliability, indicator redundancy, significance of indicator weights, indicator effect sizes, Simpsons' paradox instances associated with indicators, model-wide factor redundancy, and use of analytic composites. While some of these recommendations have been made before, others are new. Moreover, this is to our knowledge the first time that recommendations are anchored on a new measurement residual theory, namely MRT, which is a novel contribution made in this paper.

6.1. No causal arrows pointing in or out

Consistently with our discussion of MRT, we view factors as corresponding to mental ideas that must exist in the mind of a researcher before the person devises questions or question-statements to measure them. Therefore, from a factor-indicator causal direction perspective, factors *always* cause indicators, whether measurement is reflective (one-dimensional) or formative (multi-dimensional). The cause must always exist before the effect. The notion that factors cause indicators is usually represented via arrows pointing out from factors to their corresponding indicators.

At the same time, as we have seen before in our MRT formulation, a factor can be seen as an aggregation of indicators and measurement residual, where the measurement residual is akin to an "extra indicator" that is uncorrelated with the actual indicators associated with the factor. This aggregation type of relationship is often thought of as one with arrows pointing in, from the indicators and measurement residual, to the factor. But, since factors *always* cause indicators, whether measurement is reflective or formative, using arrows pointing in or out to represent factor-indicator associations can be confusing. Thus, we recommend that arrows pointing in or out should generally be avoided outside schematic representations (as we do in our discussion of MRT), particularly in empirical studies reporting results of analyses employing indicators.

6.2. Acceptable factor reliability

From MRT we know that the measurement residual weight equals the positive square root of the complement of the true reliability of a factor: $\omega_{i\varepsilon} = \sqrt{1 - \rho_i}$. Hence, it follows that the true reliability equals the complement of the measurement residual weight squared, or $\rho_i = 1 - \omega_{i\varepsilon}^2$. The measurement residual weight squared ($\omega_{i\varepsilon}^2$) equals the percentage of explained variance in the factor that is accounted for by the measurement residual. An important MRT assertion is that the measurement residual is added to the true composite to yield the corresponding factor, and thus contributes to widen the spread of the factor scores away from the true values.

We should expect the measurement residual to account for a relatively low amount of the variance in the factor, with most of that variance being accounted for by the true composite. Arguably this should be lower than .32 (or 32 percent), which is the amount of variance explained in the factor that is outside the region comprising -1 and 1 standard deviations from the true factor scores. Falling in these outer regions (left and right) is often associated with the idea of significant difference from the true values (Kock, 2016; Kock & Hadaya, 2018; Miller & Wichern, 1977; Spatz, 2010). Therefore, it is reasonable to expect the quantity ${\omega_{i\varepsilon}}^2$ to be lower than .32, which means a true factor reliability (ρ_i) of .68 or greater. In other words, true reliabilities for factors should be .68 or greater, as a condition for acceptable formative measurement quality. It is interesting to note that, given MRT's formulation, this also applies to reflective latent variables; in fact, this recommendation is in line with reliability recommendations for reflective measurement (Kline, 2010).

As noted earlier, fairly accurate estimates of the true reliabilities for factors can be obtained via maximum likelihood confirmatory factor analysis (Kline, 2010; Mueller, 1996), and the consistent partial least squares technique (Dijkstra & Schermelleh-Engel, 2014; Kock, 2019). In our empirical illustration, discussed later in this paper, true reliability estimation employs the consistent partial least squares technique (Kock, 2019).

6.3. Low indicator redundancy

As illustrated in our discussion of MRT, formative measurement attempts to capture different dimensions of the same mental idea, which itself is measured by a factor. If the dimensions captured are really conceptually different from one another, they should ideally be completely non-redundant and thus uncorrelated. However, this goal is theoretical and typically not achievable with actual populations, which are finite. In finite populations, nonzero correlations among formative indicators are to be expected, even due to spurious influences, so a certain amount of redundancy may occur. However, this redundancy should be below an acceptable level (Cenfetelli & Bassellier, 2009; Kock, 2014; Kock & Lynn, 2012; Petter et al., 2007).

Full collinearity variance inflation factors have been proposed as a measure of redundancy, with values of 3.3 or below being an indication of no significant redundancy among observed variables (Kock & Lynn, 2012). This same criterion has also been demonstrated to capture pathological common variation, through a common method bias test relying on full collinearity variance inflation factors (Kock, 2015a). Full collinearity variance inflation factors applied to indicators can be easily calculated through Eq. 6.

$$VIF_i = 1/(1 - R_i^2). \qquad \text{(Eq. 6)}$$

In Eq. 6, VIF_i is the full collinearity variance inflation factor for an indicator indexed by i of a factor, and R_i^2 is the percentage of explained variance in that indicator by the other indicators that are associated with the same factor. Unlike the correlation coefficient, which can be seen as a measure of redundancy between two indicators, the full collinearity variance inflation factor for each indicator is a measure of redundancy among each indicator and all of the other indicators associated with the same factor. Consistently with the foregoing discussion, we suggest that all full collinearity variance inflation factors for the indicators of a factor should be 3.3 or lower, as a condition for acceptable formative measurement quality.

6.4. Statistically significant weights

In MRT a factor is mathematically defined as a weighted sum of its indicators and measurement residual. Since we expect the measurement residual to be relatively low, it follows that we also expect the true reliability associated with a factor to be relatively high. Given MRT's formulation, where a factor's true reliability is the variance explained in the factor by the indicators, this would consequently lead to the expectation that each of the indicator weights be significantly different from zero. If an indicator weight is close to zero, then the indicator in question would typically make a negligible contribution to the explained variance in the corresponding factor.

Whether an indicator weight is significantly different from zero can be ascertained via the calculation of the confidence interval with left and right limit values of $\omega - z\sigma$ and $\omega + z\sigma$, where: ω is the indicator weight, z is the z-score associated with the sum of the confidence level and half of the significant level chosen (e.g., $.95 + .05/2 = .975$; $z_{.975} = 1.96$), and σ is the standard error associated with the indicator weight. If the confidence interval does not contain the number zero, we can then conclude that the indicator weight is significantly different from zero. Alternatively, one can ascertain whether an indicator weight is significantly different from zero by calculating a one-tailed P value associated with the weight, and checking if that P value is lower than a generally accepted threshold (usually .05); the P value being lower would suggest that the weight is significantly different from zero. One way of another, all indicator weights are expected to be significantly different from zero in a statistical sense, as a condition for acceptable formative measurement quality.

6.5. Acceptable indicator effect sizes

From MRT we derive the expectation that indicator weights be significantly different from zero. But statistical significance tests are highly sensitive to sample size, regardless of whether they are based on confidence intervals or P values. This is due to the fact that standard errors decrease with sample size, and is true also for statistical significance tests applied to indicator weights. The larger the size of the sample being analyzed, the more likely it is that a small indicator weight will be found to be significantly different from zero in a statistical sense. This can lead to type I errors, where indicator weights are found to be nonzero when, in fact, they are zero at the population level.

Effect sizes (Cohen, 1988; 1992; Kock & Hadaya, 2018) are calculated to avoid the above-mentioned shortcoming of statistical significance tests. In the context of multivariate least squares regression, the most widely used measure of effect size is Cohen's *f*-squared coefficient (Cohen, 1988; 1992), which is calculated though Eq. 7. In Eq. 7, f_{ij}^2 is Cohen's *f*-squared coefficient for the indicator j of factor i, ΔR_{ij}^2 is the incremental contribution of the indicator j to the percentage of explained variance in the factor i, and R_i^2 is the total percentage of explained variance in the factor i by all of the indicators (or the true reliability associated with the factor). Another measure of effect size that has been frequently used in the past is simply ΔR_{ij}^2, which is usually slightly lower than Cohen's *f*-squared coefficient, and is thus seen as leading to more conservative assessments of effect size (Kock & Hadaya, 2018).

$$f_{ij}^2 = \Delta R_{ij}^2 / (1 - R_i^2). \qquad \text{(Eq. 7)}$$

By convention, effect sizes of .02, .15, and .35 are respectively termed small, medium, and large; and effect sizes below .02 are considered to be too small to be associated with non-negligible effects (Cohen, 1992; Kock & Hadaya, 2018). For example, a rather small indicator weight of .05 would be found to be statistically significant with a sample size of 5,000, whether we used a confidence interval or P value to assess significance. However, the effect size associated with the corresponding indicator (estimated through Cohen's *f*-squared coefficient) would be approximately .008, assuming a reliability of .68 (the minimum acceptable reliability). This would be well below .02 and thus too small to be non-negligible. As a general rule consistent with the foregoing discussion, we argue that indicator effect sizes should be .02 or greater, as a condition for acceptable formative measurement quality.

6.6. No Simpsons' paradox instances

As mentioned before, in MRT the true reliability associated with a factor is defined as the amount of variance explained in the factor by the corresponding indicators. Normally each indicator is expected to contribute positively to the variance explained in the factor. However, this is not the case when an instance of Simpsons' paradox (Pavlides & Perlman, 2009; Wagner, 1982) occurs for a given indicator of a factor. When Simpsons' paradox occurs in this context, the indicator in question makes a *negative* individual contribution to the variance explained in the factor (Kock & Gaskins, 2016; Pearl, 2009).

An easy way to identify an indicator associated with an instance of Simpsons' paradox is to inspect the weights and loadings associated with the indicator. Simpsons' paradox is characterized by an indicator's weight and loading presenting different signs (Kock & Gaskins, 2016). For example, a loading may be positive and the weight negative, or vice-versa. In such a case, the negative contribution to the variance explained in the factor suggests that measurement model misspecification occurred; i.e., the possibility that the indicator does not actually belong to the factor, in a manner of speaking. Given this, we argue that no Simpsons' paradox instances should exist in connection with any indicator of a factor, as a condition for acceptable formative measurement quality.

6.7. Low model-wide factor redundancy

Within the framework put forth by MRT, formative measurement aims to capture different dimensions of the same mental idea, which is itself measured by a factor. In this context it is important to ensure that the indicators used in formative measurement capture different dimensions of the same mental idea, and not of different mental ideas that are measured by different factors in the same model. In other words, it is important to ensure that indicators used in formative measurement capture variation from their own factor, and not from other factors. If correlations exist among indicators of one formative factor and other factors, those correlations should be entirely due to the network of causal effects linking the factors. They should not be due to the indicators associated with a factor mistakenly measuring other factors.

Full collinearity variance inflation factors applied to factors can be used to test for the above, in a way that is similar to their use at the indicator level (discussed earlier). Full collinearity variance inflation factors have been proposed as a measure of model-wide pathological redundancy among factors, with values of 5 or below being an indication of no significant redundancy (Kock & Lynn, 2012; Kock, 2015a). A recent study suggests the threshold of 10 (instead of 5) in models where all latent variables are modeled as factors and none as composites (Kock & Dow, forthcoming).

These full collinearity variance inflation factors would be calculated for each factor indexed by i as $VIF_i = 1/(1 - R_i^2)$, where R_i^2 is the percentage of explained variance in the factor by all of the other factors in the model. In line with this discussion, we suggest that all full collinearity variance inflation factors for the factors in a model should be 5 or lower. Alternately, the more relaxed rule that they should be 10 or lower (see: Kock & Dow, forthcoming), could be used as a condition for acceptable formative measurement quality.

6.8. Use of analytic composites

Based on conceptual foundation provided by MRT we can assume that low indicator redundancy is a desirable characteristic in formative latent variable measurement. However, this leads to low correlations among indicators and thus low correlations among indicators and their factors. Consequently, whenever low indicator redundancy occurs, factor reliabilities will be low unless many indicators are used for formative measurement. However, low indicator redundancy leads to low loadings and weights, particularly if many indicators are used, with the values of the weights approaching those of the loadings. In fact, zero correlations among indicators, or no redundancy, would lead to weights and loadings of exactly the same magnitude. In this scenario, low weights would tend to be non-significant and associated with very small effect sizes.

The competing trends above can be balanced by the use of analytic composites to capture the variation from indicators, using the same weights as in the original analysis (prior to the creation of the analytic composites), for those indicators that individually do not pass tests of acceptable formative measurement quality. Such analytic composites will, by definition, incorporate variation only from the indicators used; i.e., they will be akin to instrumental variables (Arellano & Bover, 1995; Kline, 2010). Using analytic composites in this way should ensure that the reliability associated with factors will reach acceptable levels, as long as the analytic composites themselves meet acceptable formative measurement quality criteria. This will in turn happen if the analytic composites aggregate indicators that are indeed formatively associated with the same latent variable.

Therefore, we propose that analytic composites should be used to aggregate indicators that individually do not meet acceptable formative measurement quality criteria, using the original weights, as long as the analytic composites themselves meet those criteria. Normally the addition or removal of an indicator to or from an analytic composite will affect whether the analytic composite meets the criteria. This property can be used to define what indicators end up being part of the analytic composite, what indicators remain separate, and what indicators are completely excluded from the factor.

7. Empirical illustration

In this section, and component sub-sections, we illustrate formative measurement in light of our proposed MRT and related model revisions to address the related recommendations aimed at promoting acceptable formative measurement quality. Our illustration is based on data from 290 team-based product innovation efforts conducted in organizations in the Northeastern U.S. These team efforts led to the creation of new products, or major redesign of existing products; comprising manufacturing goods, services, and information products. Examples of such products are car parts, airplane engines, courses about academic topics, indices used for organizational planning, and software; among others.

7.1. The illustrative model

The illustrative model used for analyses is shown in Figure 3. This model has been developed based on media naturalness theory (DeRosa et al., 2004; Kock, 2004; 2005; Serrano & Karahanna, 2016), from which various empirical studies have been derived and validations conducted (Akgün et al., 2014; Kock & Lynn, 2012; Kock et al., 2006; Serrano & Karahanna, 2016; Zahedi et al., 2016). The model contains four latent variables: *electronic communication media use (ECM)* by each product innovation team, measured formatively through 16 indicators; *project management techniques use (Prjmgt)* by each team, measured reflectively through 3 indicators; *product innovation efficiency (Effic)* of each team, measured reflectively through 5 indicators; and *new product success (Success)* of each team, measured reflectively through 7 indicators.

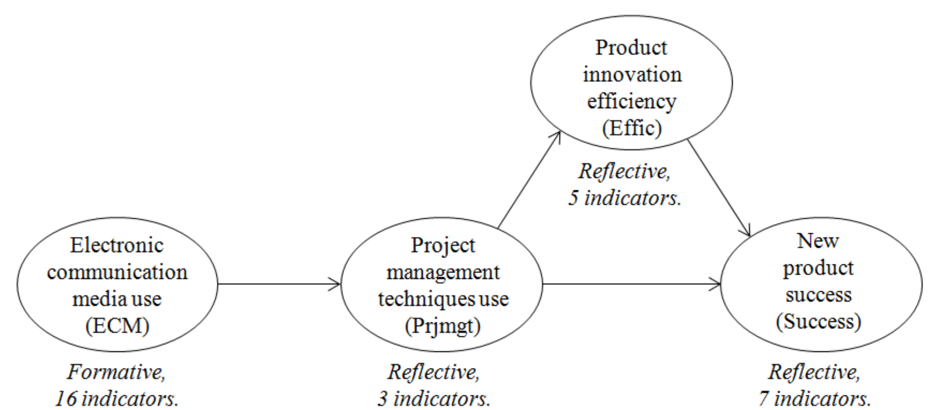

Figure 3. Illustrative model

The decisions as to whether each of the latent variables were treated as formative or reflective were made after the data was collected and analyzed, using the loadings rule discussed earlier, and were consistent with theoretical expectations at the design stage of the question-statements. Appendix C provides a list of the question-statements associated with these latent variables and corresponding indicators. All of the reflective latent variables passed widely used measurement assessment tests addressing convergent validity, discriminant validity, and reliability (Kline, 2010; Kock & Lynn, 2012).

While our goal in this paper is not to test the model empirically, but to use it for illustration purposes, it is useful for readers to know that the model incorporates the following beliefs in the context of geographically dispersed product innovation teams. *Electronic communication media use (ECM)* is believed to facilitate and thus cause an increase in *project management techniques use (Prjmgt)*, which in turn mediates its (i.e., *ECM*'s) positive impact on *product innovation efficiency (Effic)* and *new product success (Success)*. An important element of the model is the belief that technology use (i.e., *ECM*) does not exert its positive effects, with geographically dispersed product innovation teams, without the concomitant use of project management techniques. In other words, technology alone is not of much help to product innovation teams without the use of project management techniques.

Our analyses were conducted with the software WarpPLS 6.0 (Kock, 2018), because this software conveniently estimates true composites and factors through its "Factor-Based PLS Type CFM3" algorithm in a way that is fully compatible with our MRT formulation and equations (Kock, 2015b; 2017; Kock & Sexton, 2017). The free trial version of this software is a full implementation (not a demo version) and is available for approximately 3 months. Moreover, this software estimates all of the coefficients needed for formative measurement model assessment: true factor reliabilities, full collinearity variance inflation factors for indicators and latent variables, true weights and respective P values, indicator effect sizes, and weight-loading signs for Simpsons' paradox identification. Finally, this software provides a specialized feature that allows for the creation of analytic composites. The software features employed yielded intermediate and final results that were checked with other widely used software tools such as

SPSS, MATLAB, and various R packages. These checks often had to rely on extensive manual work, and generally suggested that the features yield trustworthy results.

7.2. Results of analyses before and after revision

Tables 1 and 2 show the measurement quality assessment results prior to formative variable revision, and after revision, respectively. The revision involved the use of an analytic composite to aggregate indicators that did not individually meet formative measurement quality criteria.

Table 1. Measurement quality assessment results prior to formative variable revision

Indicator weights, P values, FCVIFs, WLSs and ESs					
	Weight	P value	FCVIF	WLS	ES
ECM1	.195	<.001	4.434	1	.047
ECM2	.055	.171	4.693	1	.024
ECM3	.124	.016	1.217	1	.070
ECM4	.129	.013	1.344	1	.029
ECM5	.137	.009	1.327	1	.084
ECM6	.166	.002	1.395	1	.061
ECM7	.157	.003	1.392	1	.102
ECM8	.101	.040	1.445	1	.028
ECM9	.154	.004	3.635	1	.040
ECM10	.036	.270	3.496	1	.010
ECM11	.125	.016	1.406	1	.000
ECM12	.072	.108	1.472	1	.045
ECM13	.087	.067	1.410	1	.032
ECM14	.091	.059	1.388	1	.010
ECM15	.031	.298	1.632	1	.004
ECM16	.096	.048	1.713	1	.004
Model-wide latent variable FCVIFs and reliabilities					
	ECM	Prjmgt	Effic	Success	
FCVIF	1.092	1.525	1.812	1.565	
FR	.780	.800	.897	.966	

Notes: FCVIF = full collinearity variance inflation factor; WLS = weight-loading sign (-1 = weight and loading with different signs); ES = effect size; FR = true reliability associated with factor; shaded rows refer to indicators that did not pass measurement quality criteria.

The acronym FCVIF refers to full collinearity variance inflation factor. All FCVIFs are shown for indicators and factors. WLS refers to weight-loading sign (-1 means that the weight and loading have different signs); ES refers to indicator effect size; and FR refers to the true reliability associated with factor. The shaded rows in Table 1 refer to indicators that did not pass one or more of the measurement quality criteria discussed earlier.

Table 2. Measurement quality assessment results after formative variable revision					
Indicator weights, P values, FCVIFs, WLSs and ESs					
	Weight	P value	FCVIF	WLS	ES
ECM3	.178	<.001	1.183	1	.105
ECM4	.162	.003	1.287	1	.039
ECM5	.267	<.001	1.387	1	.172
ECM6	.189	<.001	1.310	1	.073
ECM7	.217	<.001	1.349	1	.148
ECM8	.133	.011	1.369	1	.039
AC(1-2,9-16)	.228	<.001	1.713	1	.130
Model-wide latent variable FCVIFs and reliabilities					
	ECM	Prjmgt	Effic	Success	
FCVIF	1.126	1.551	1.809	1.569	
FR	.728	.800	.897	.966	
Notes: the analytic composite AC(1-2,9-16) aggregates the indicators that did not pass measurement quality criteria (shaded cells in previous table); after revision (this table), all indicators and analytic composite pass measurement quality criteria.					

The analytic composite AC(1-2,9-16), shown in Table 2, aggregates the indicators that did not pass measurement quality criteria. The indicators were aggregated initially based on their original weights; these changed after standardization, in terms of their absolute values, but retained the same relative weight proportions against one another. The criteria violations are indicated in shaded cells in the previous table. These indicators are ECM1, ECM2, and ECM9 ... ECM16. After the revision leading to the removal of the offending indicators as standalone indicators and their aggregation into the analytic composite, all remaining indicators (i.e., ECM3 … ECM8) and analytic composite AC(1-2,9-16) passed measurement quality criteria.

It is noteworthy that the path coefficient for the link *ECM* → *Prjmgt* was .265 prior to formative variable revision, and .313 after revision. Since the model arguably has a sound theoretical basis, this increase in path coefficient strength is consistent with the idea that the revision led to the use of a better combination of indicators and weights with respect to formative measurement.

It is also noteworthy that if we removed the analytic composite from the set of indicators, this path coefficient would display a slight increase to .320; but its factor reliability would go down to .670, which is below the recommended threshold of .680 for acceptable factor reliability. In other words, the slight increase from .313 to .320 may "look good" from a theoretical standpoint but might in fact be due to spurious capitalization on error, since it is associated with an increase in the magnitude of the measurement residual. This possible capitalization on error is prevented by the combined application of the formative measurement quality assessment criteria.

7.3. What if we use PLS Mode B?

Researchers who subscribe to the use of classic (as opposed to factor-based, see: Kock, 2015b; 2019; Kock & Sexton, 2017) partial least squares techniques for analyses of the type discussed here may be tempted to use the algorithm known as Mode B, also known as the "formative mode" (Lohmöller, 1989), to assess formative measurement quality. The problematic nature of Mode B has been demonstrated before in various contexts (Aguirre-Urreta & Marakas, 2013; Kock & Mayfield, 2015).

Further, since Mode B, like all classic partial least squares techniques, generates composites, it is clearly incompatible with our proposed formulation of factors in MRT. It should be noted that our analyses discussed above do not employ classic partial least squares techniques (for a comprehensive discussion, see: Lohmöller, 1989), but rather factor-based techniques that explicitly model measurement error building on reliabilities obtained through the consistent partial least squares procedure (Dijkstra & Schermelleh-Engel, 2014; Kock, 2015b; Kock 2018; Kock & Sexton, 2017).

In line with the views above, critical of Mode B, our analyses discussed below suggest that this classic partial least squares mode generates an inordinate number of Simpson paradox instances at the indicator level. These analyses employed Mode B for the formative latent variable and Mode A for the reflective latent variables, as well as the path weighting scheme (Adelman & Lohmoller, 1994; Kock & Sexton, 2017). Mode A is also known as the "reflective mode" in classic partial least squares parlance (Lohmöller, 1989).

This problem with Mode B is illustrated in Table 3, where shaded cells indicate the occurrence of such instances of Simpson's paradox. After multiple attempts on our part, it seems that if we use Mode B it is virtually impossible for us to revise our formative latent variable to meet our recommended measurement quality criteria. This pattern of Simpson's paradox occurrence is virtually identical to that found by Kock & Mayfield (2015).

Table 3. Many Simpson's paradox instances when Mode B is used

Indicator weights, P values, FCVIFs, WLSs and ESs					
	Weight	P value	VIF	WLS	ES
ECM1	-.321	<.001	4.434	-1	.068
ECM2	.405	<.001	4.693	1	.156
ECM3	.309	<.001	1.217	1	.154
ECM4	-.113	.026	1.344	-1	.023
ECM5	.402	<.001	1.327	1	.218
ECM6	-.056	.168	1.395	-1	.018
ECM7	.371	<.001	1.392	1	.211
ECM8	.024	.338	1.445	1	.006
ECM9	-.104	.036	3.635	-1	.023
ECM10	.175	.001	3.496	1	.042
ECM11	-.322	<.001	1.406	-1	.001
ECM12	.573	<.001	1.472	1	.313
ECM13	.157	.003	1.410	1	.051
ECM14	-.136	.009	1.388	-1	.013
ECM15	.012	.422	1.632	1	.001
ECM16	-.185	.048	1.713	-1	.006

As mentioned before, each indicator is expected to contribute positively to the variance explained in the factor to which it is theoretically associated at the questionnaire design stage. When an instance of Simpsons' paradox occurs for a given indicator of a factor, this means that the indicator in question makes a *negative* individual contribution. This suggests that Mode B calculates factor scores in a way that inflates the contribution of some indicators to the variance explained in their factor, to make up for the negative contributions of other indicators. This likely leads to weights that defy theoretical expectations held by the questionnaire designer (e.g., the various negative weights), and possibly also defy commonsense assumptions about the phenomena being investigated.

8. Discussion

When formative measurement is used, the indicators are often viewed as causing the latent construct to which they are associated. Since the factor associated with the latent construct quantifies it, in formative measurement the indicators are seen as causing the factor, in a mathematical sense. This view is paradoxical, because the mental idea measured by the factor must exist before the question-statements that are quantified through the indicators are developed. In part due to this, and related problems, formative measurement has been heavily criticized in the past. Our perspective, discussed in this paper, is that formative measurement can indeed be used, as long as it is properly conceptualized and certain precautions are taken.

Anchored on a new measurement residual theory (acronym MRT), of which key elements were proposed in this paper, we argued that the indicators-to-factor direction of causal links normally associated with formative measurement is misguided. We also proposed a set of recommendations for the assessment of formative measurement quality, addressing: factor reliability, indicator redundancy, significance of indicator weights, indicator effect sizes, Simpsons' paradox instances associated with indicators, model-wide factor redundancy, and use of analytic composites. These are summarized in Table 4.

Table 4. Assessment of formative measurement quality	
Goal	**Recommendation**
Ensure construct-indicator semantic coherence.	All question-statements associated with the same latent construct should refer to the same unit of analysis. Indicators associated with question-statements that fail this semantic assessment check should be either removed from the analysis or re-assigned.
Determine whether to treat a latent variable as formative.	After any needed indicator removals or re-assignments, latent variables for which all indicators have loadings equal to or greater than .5 should be treated as reflective. If that is not the case, then they should be treated as formative, following the recommendations below. (If the loadings are lower than .5 for reasons such as response error or bad sampling, which are unrelated to formative measurement, these problems will arguably be uncovered by the combined use of the recommendations below.)
Avoid confusion about factor-indicator causality.	Causal factor-indicator arrows pointing in or out of factors should generally be avoided, outside schematic representations (e.g., regressand-regressor diagrams).
Ensure acceptable factor reliability.	True reliabilities for factors should be .68 or greater.
Ensure low indicator redundancy.	All full collinearity variance inflation factors for the indicators of a factor should be 3.3 or lower.
Ensure statistically significant factor-indicator associations.	All indicator weights should be significantly different from zero in a statistical sense.
Avoid excessively weak factor-indicator associations.	Indicator effect sizes should be .02 or greater, even if the corresponding weights are significantly different from zero in a statistical sense.
Foster interpretational clarity.	No Simpsons' paradox instances should exist in connection with any indicator of a factor.
Ensure low factor redundancy.	All full collinearity variance inflation factors for the factors in a model should be 5 or lower. Alternatively, the more relaxed rule that they should be 10 or lower (see: Kock & Dow, forthcoming) can be used.
Avoid loss of variation from non-conforming indicators.	Analytic composites should be used to aggregate indicators that individually do not meet acceptable formative measurement quality criteria, using the original weights, as long as the analytic composites themselves meet those criteria.

Our new conceptualization and related theoretical elements refer primarily to measurement error arising from the use of questionnaires. It incorporates key elements from two fairly well-established theoretical frameworks. These frameworks form the foundation of latent variable measurement and structural equation modeling, but to our knowledge have never had before their key elements integrated into a single theory. The frameworks in question are classic measurement error theory (Nunnaly, 1978; Nunnally & Bernstein, 1994) and the common factor model (Kline, 2010; MacCallum & Tucker, 1991).

Additionally, we defined analytic composites as aggregations of indicators without measurement error, which makes them different from factors. We argued that the weights of analytic composites should be defined by their designer; based on the results of factor-based analyses, or prior theory and related research. It is important to stress that, in our definition, we noted that one common characteristic of analytic composites is that they are *designed* to serve a purpose, providing as examples the Standard & Poor's 500 and the Dow Jones Industrial Average, which are stock market indices; and the Gini coefficient, a country wealth distribution index. We also showed that analytic composites can be used to aggregate indicators that do not individually pass formative measurement quality tests. We used the term analytic composite instead of index to avoid confusion with the terminology normally used in structural equation modeling, where the word index is frequently used to refer to measures of model fit.

9. Conclusion

In this paper, formative measurement was illustrated in light of our proposed measurement residual theory, and related model revisions needed to address the recommendations to ensure acceptable formative measurement quality. The illustrative study employed for this was based on data from 290 teams who conducted product innovation projects in organizations in the Northeastern U.S. The teams used various electronic media to communicate. The projects conducted by the teams led to the creation of new products, or major redesign of existing products; which included manufacturing goods, services, and information products.

We demonstrated the problematic nature of the use the partial least squares algorithm known as Mode B, also branded as the “formative mode” (Lohmöller, 1989), when it is used to estimate models with formative latent variables. Our analyses showed that Mode B generates an inordinate number of Simpson paradox instances at the indicator level. When an instance of

Simpsons' paradox occurs for a given indicator of a factor, the indicator in question makes a *negative* individual contribution to the explained variance in the factor; possibly defying theoretical expectations held by the questionnaire designer and commonsense assumptions about the phenomena being investigated.

Our perspective and related recommendations differ significantly from previous guidelines provided in the context of partial least squares and related composite-based methods. They also differ from guidelines provided in the context of maximum likelihood and related methods (Bollen, 2011; Diamantopoulos, 2011). Formative measurement implementation and assessment in structural equation modeling employing maximum likelihood and related methods typically rely on modeling indicators as being part of the structural model, as opposed to being part of the measurement model (Bollen, 2011). As such, problematic identification issues frequently emerge (Diamantopoulos, 2011; Hsu et al., 2018). Our proposed solutions for formative measurement implementation and assessment are not expected to make identification problems worse, but rather to ameliorate them.

This paper makes theoretical and methodological contributions that we hope will help researchers implement formative measurement in ways that are scholarly defensible from a measurement quality perspective. We also hope that the new theoretical perspective and related measurement quality assessment recommendations will take the ongoing debate on formative measurement one step further.

References

Adelman, I., & Lohmoller, J.-B. (1994). Institutions and development in the nineteenth century: A latent variable regression model. *Structural Change and Economic Dynamics*, 5(2), 329-359.

Aguirre-Urreta, M. I., & Marakas, G. M. (2013). Partial least squares and models with formatively specified endogenous constructs: A cautionary note. *Information Systems Research*, 25(4), 761-778.

Akgün, A. E., Lynn, G. S., Keskin, H., & Dogan, D. (2014). Team learning in IT implementation projects: Antecedents and consequences. *International Journal of Information Management*, 34(1), 37-47.

Ambalov, I. A. (2021). An investigation of technology trust and habit in IT use continuance: a study of a social network. *Journal of Systems and Information Technology*, 23(1), 53-81.

Amora, J. T. (2023). On the validity assessment of formative measurement models in PLS-SEM. *Data Analysis Perspectives Journal*, 4(2), 1-7.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29-51.

Bag, S., Srivastava, G., Bashir, M. M. A., Kumari, S., Giannakis, M., & Chowdhury, A. H. (2022). Journey of customers in this digital era: Understanding the role of artificial intelligence technologies in user engagement and conversion. *Benchmarking: An International Journal*, 29(7), 2074-2098.

Bartlett, M.S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28(1), 97-104.

Bayonne, E., Marin-Garcia, J. A., & Alfalla-Luque, R. (2020). Partial least squares (PLS) in operations management research: Insights from a systematic literature review. *Journal of Industrial Engineering and Management*, 13(3), 565-597.

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, 35(2), 359-372.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305-314.

Canatay, A., Emegwa, T., Lybolt, L. M., & Loch, K. D. (2022). Reliability assessment in SEM models with composites and factors: A modern perspective. *Data Analysis Perspectives Journal*, 3(1), 1-6.

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, 33(4), 689-707.

Cheah, J. H., Ting, H., Ramayah, T., Memon, M. A., Cham, T. H., & Ciavolino, E. (2019). A comparison of five reflective–formative estimation approaches: reconsideration and recommendations for tourism research. *Quality & Quantity*, 53(3), 1421-1458.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250-1262.

DeRosa, D. M., Hantula, D. A., Kock, N., & D'Arcy, J. (2004). Trust and leadership in virtual teamwork: A media naturalness perspective. *Human Resource Management*, 43(2-3), 219-232.
Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, 35(2), 335-358.
Dijkstra, T.K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, 79(4), 585-604.
Dillon, W.R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York, NY: Wiley.
DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20), 1-11.
Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, 14(2), 370-388.
Fleuren, B. P., van Amelsvoort, L. G., Zijlstra, F. R., de Grip, A., & Kant, I. (2018). Handling the reflective-formative measurement conundrum: A practical illustration based on sustainable employability. *Journal of Clinical Epidemiology*, 103(1), 71-81.
Giovanis, A., Pierrakos, G., Rizomyliotis, I., & Binioris, S. (2018). Assessing service quality and customer behavioral responses in hospital outpatient services: A formative measurement approach. *International Journal of Quality and Service Sciences*, 10(2), 98-116.
Hanafiah, M. H. (2020). Formative vs. reflective measurement model: Guidelines for structural equation modeling research. *International Journal of Analysis and Applications*, 18(5), 876-889.
Hardin, A. (2017). A call for theory to support the use of causal-formative indicators: A commentary on Bollen and Diamantopoulos. *Psychological Methods*, 22(3), 597-604.
Hardin, A., & Marcoulides, G. A. (2011). A commentary on the use of formative measurement. *Educational and Psychological Measurement*, 71(5), 753-764.
Hardin, A., Chang, J. C. J., Fuller, M. A., & Torkzadeh, G. (2011). Formative measurement and academic research: In search of measurement theory. *Educational and Psychological Measurement*, 71(2), 281-305.
Ho, C. I., Chen, M. C., & Shih, Y. W. (2022). Customer engagement behaviours in a social media context revisited: Using both the formative measurement model and text mining techniques. *Journal of Marketing Management*, 38(7-8), 740-770.
Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205-218.
Howell, R. D., Breivik, E., & Wilcox, J. B. (2007b). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods*, 12(2), 238-245.
Howell, R. D., Breivik, E., & Wilcox, J. B. (2013). Formative measurement: A critical perspective. *Data Base for Advances in Information Systems*, 44(4), 44-55.
Hsu, J., Chiu, C. M., Lowry, P. B., & Liang, T. P. (2018). Solving the interpretational-confounding and interpretational-ambiguity problems of formative construct modeling in behavioral research: Proposing a two-stage fixed-weight redundancy approach. *Journal of the Association for Information Systems*, 19(7), 618-671.
Kim, G., Shin, B., & Grover, V. (2010). Investigating two contradictory views of formative measurement in information systems research. *MIS Quarterly*, 34(2), 345-365.

Kline, R.B. (2010). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.

Kock, N. (2004). The psychobiological model: Towards a new theory of computer-mediated communication based on Darwinian evolution. *Organization Science*, 15(3), 327-348.

Kock, N. (2005). Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. *IEEE Transactions on Professional Communication*, 48(2), 117-130.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(3), 1-13.

Kock, N. (2015a). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1-10.

Kock, N. (2015b). A note on how to conduct a factor-based PLS-SEM analysis. *International Journal of e-Collaboration*, 11(3), 1-9.

Kock, N. (2016). Hypothesis testing with confidence intervals and P values in PLS-SEM. *International Journal of e-Collaboration*, 12(3), 1-6.

Kock, N. (2017). Structural equation modeling with factors and composites: A comparison of four methods. *International Journal of e-Collaboration*, 13(1), 1-9.

Kock, N. (2018). *WarpPLS User Manual: Version 6.0*. Laredo, TX: ScriptWarp Systems.

Kock, N. (2019). From composites to factors: Bridging the gap between PLS and covariance-based structural equation modeling. *Information Systems Journal*, 29(3), 674-706.

Kock, N. (2021a). Common structural variation reduction in PLS-SEM: Replacement analytic composites and the one fourth rule. *Data Analysis Perspectives Journal*, 2(5), 1-6.

Kock, N. (2021b). Moderated mediation and J-curve emergence in path models: An information systems research perspective. *Journal of Systems and Information Technology*, 23(3), 303-321.

Kock, N. (2022). Using causality assessment indices in PLS-SEM. *Data Analysis Perspectives Journal*, 3(5), 1-6.

Kock, N. (2023). Contributing to the success of PLS in SEM: An action research perspective. *Communications of the Association for Information Systems*, 52(1), 730-734.

Kock, N., & Dow, K.E. (forthcoming). Statistical significance and effect size tests in SEM: Common method bias and strong theorizing. *Advances in Management Accounting*.

Kock, N., & Gaskins, L. (2016). Simpson's paradox, moderation, and the emergence of quadratic relationships in path models: An information systems illustration. *International Journal of Applied Nonlinear Science*, 2(3), 200-234.

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261.

Kock, N., & Lynn, G.S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546-580.

Kock, N., & Mayfield, M. (2015). PLS-based SEM algorithms: The good neighbor assumption, collinearity, and nonlinearity. *Information Management and Business Review,* 7(2), 113-130.

Kock, N., & Moqbel, M. (2021). Social networking site use, positive emotions, and job performance. *Journal of Computer Information Systems*, 61(2), 163-173.

Kock, N., & Sexton, S. (2017). Variation sharing: A novel numeric solution to the path bias underestimation problem of PLS-based SEM. *International Journal of Strategic Decision Sciences*, 8(4), 46-68.
Kock, N., Lynn, G. S., Dow, K. E., & Akgün, A. E. (2006). Team adaptation to electronic communication media: evidence of compensatory adaptation in new product development teams. *European Journal of Information Systems*, 15(3), 331-341.
Kono, S., Ito, E., & Loucks-Atkinson, A. (2021). Are leisure constraints models reflective or formative? Evidence from confirmatory tetrad analyses. *Leisure Sciences*, 44(1), 55-76.
Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg, Germany: Physica-Verlag.
MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, 109(3), 502-511.
Mamakou, X. J., Cohen, S., & Manolopoulos, D. (2024). Post-implementation evaluation of enterprise resource planning (ERP) systems: an internal auditors' perspective. *Journal of Systems and Information Technology*, 26(3), 363-394.
McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on partial least squares path modeling. *Organizational Research Methods*, 7(2), 210-251.
Mikulić, J., & Ryan, C. (2018). Reflective versus formative confusion in SEM based tourism research: A critical comment. *Tourism Management*, 68(1), 465-469.
Miller, R.B., & Wichern, D.W. (1977). *Intermediate business statistics: Analysis of variance, regression and time series*. New York, NY: Holt, Rihehart and Winston.
Mohammad Salameh, A. A., Ahmad, H., Zulhumadi, F., & Abubakar, F. M. (2018). Relationships between system quality, service quality, and customer satisfaction: M-commerce in the Jordanian context. *Journal of Systems and Information Technology*, 20(1), 73-102.
Mueller, R.O. (1996). *Basic principles of structural equation modeling*. New York, NY: Springer.
Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
Nunnaly, J.C. (1978). *Psychometric theory.* New York, NY: McGraw Hill.
Pavlides, M.G., & Perlman, M.D. (2009). How likely is Simpson's paradox? *The American Statistician*, 63(3), 226-233.
Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press.
Peštović, K., Milicevic, N., Djokic, N., & Djokic, I. (2021). Audit service quality perceived by customers: Formative modelling measurement approach. *Sustainability*, 13(21), 1-16.
Peterson, R.A., & Yeolib, K. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98(1), 194-198.
Petter, S. (2018). Haters gonna hate: PLS and information systems research. *Data Base for Advances in Information Systems*, 49(2), 10-13.
Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 31(4), 623-656.
Serrano, C. I., & Karahanna, E. (2016). The compensatory interaction between user capabilities and technology capabilities in influencing task performance: An empirical assessment in telemedicine consultations. *MIS Quarterly*, 40(3), 597-621.
Spatz, C. (2010). *Basic statistics: Tales of distributions*. Stamford, CT: Wadsworth.

Theodosiou, M., Katsikea, E., Samiee, S., & Makri, K. (2019). A comparison of formative versus reflective approaches for the measurement of electronic service quality. *Journal of Interactive Marketing*, 47(1), 53-67.

Thurstone, L.L. (1935). *The vectors of mind*. Chicago, IL: University of Chicago Press.

Wagner, C.H. (1982). Simpson's paradox in real life. *The American Statistician*, 36(1), 46–48.

Zahedi, F. M., Walia, N., & Jain, H. (2016). Augmented virtual doctor office: Theory-based design and assessment. *Journal of Management Information Systems*, 33(3), 776-808.

Appendix A: Derivation of true composite weights equation

The true composite weights ω_i (stored in a column vector) must satisfy the equation below, where x_i is a matrix where each column refers to one of the indicators associated with composite C_i (and thus with factor F_i); $\lambda_i{}'$ is the transpose of λ_i, the column vector storing the loadings associated with the indicators; θ_i is the matrix of indicator error terms; $\Sigma_{x_i x_i}$ is the covariance matrix of the indicators; and $\Sigma_{x_i\theta_i}$ is the matrix of covariances among indicators and their individual errors. The superscript -1 denotes the classic matrix inversion; the function $diag(\cdot)$ returns a matrix with only its diagonal elements different from zero; and the superscript $+$ denotes the Moore–Penrose pseudoinverse transformation.

$$\omega_i = \Sigma_{x_i x_i}{}^{-1}\left(\Sigma_{x_i x_i} - diag\left(\Sigma_{x_i\theta_i}\right)\right)\lambda_i{}'^{+}.$$

From our previous discussion on MRT, we know that

$x_i = F_i\lambda_i{}' + \theta_i,\ F_i = x_i\omega_i + \varepsilon_i\omega_{i\varepsilon}$.

Combining these two equations we obtain

$x_i = (x_i\omega_i + \varepsilon_i\omega_{i\varepsilon})\lambda_i{}' + \theta_i \rightarrow$

$x_i = x_i\omega_i\lambda_i{}' + \varepsilon_i\omega_{i\varepsilon}\lambda_i{}' + \theta_i$.

Applying covariance properties to the above we obtain

$\Sigma_{x_i x_i} = \Sigma_{x_i x_i}\omega_i\lambda_i{}' + \Sigma_{x_i\varepsilon_i}\omega_{i\varepsilon}\lambda_i{}' + \Sigma_{x_i\theta_i} \rightarrow$

$\Sigma_{x_i x_i} = \Sigma_{x_i x_i}\omega_i\lambda_i{}' + diag\left(\Sigma_{x_i\theta_i}\right) \rightarrow$

$\Sigma_{x_i x_i}\omega_i\lambda_i{}' = \Sigma_{x_i x_i} - diag\left(\Sigma_{x_i\theta_i}\right) \rightarrow$

$\omega_i\lambda_i{}' = \Sigma_{x_i x_i}{}^{-1}\left(\Sigma_{x_i x_i} - diag\left(\Sigma_{x_i\theta_i}\right)\right)$,

where the superscript -1 denotes the classic matrix inversion.

In order to isolate ω_i in the equation above, we need to use the Moore–Penrose pseudoinverse transformation, because the classic matrix inversion transformation cannot be applied to a vector. Doing this, we obtain

$\omega_i = \Sigma_{x_i x_i}{}^{-1}\left(\Sigma_{x_i x_i} - diag\left(\Sigma_{x_i\theta_i}\right)\right)\lambda_i{}'^{+}$,

where the superscript $+$ denotes the Moore–Penrose pseudoinverse transformation.

Appendix B: Creating an analytic composite in WarpPLS

We show below how one can create an analytic composite with the software WarpPLS 6.0 (Kock, 2018), because this software conveniently allows for this task to be accomplished in a way that is fully compatible with our MRT formulation and equations. Once a structural equation modeling analysis is conducted with WarpPLS, the menu option “Explore analytic composites and instrumental variables” becomes available (see Figure B.1), allowing users to create analytic composites. Readers are referred to WarpPLS.com for a 5-minute video on how to conduct a full structural equation modeling analysis. To get to this video click on “YouTube videos” and then on “SEM Analysis with WarpPLS (all steps)”.

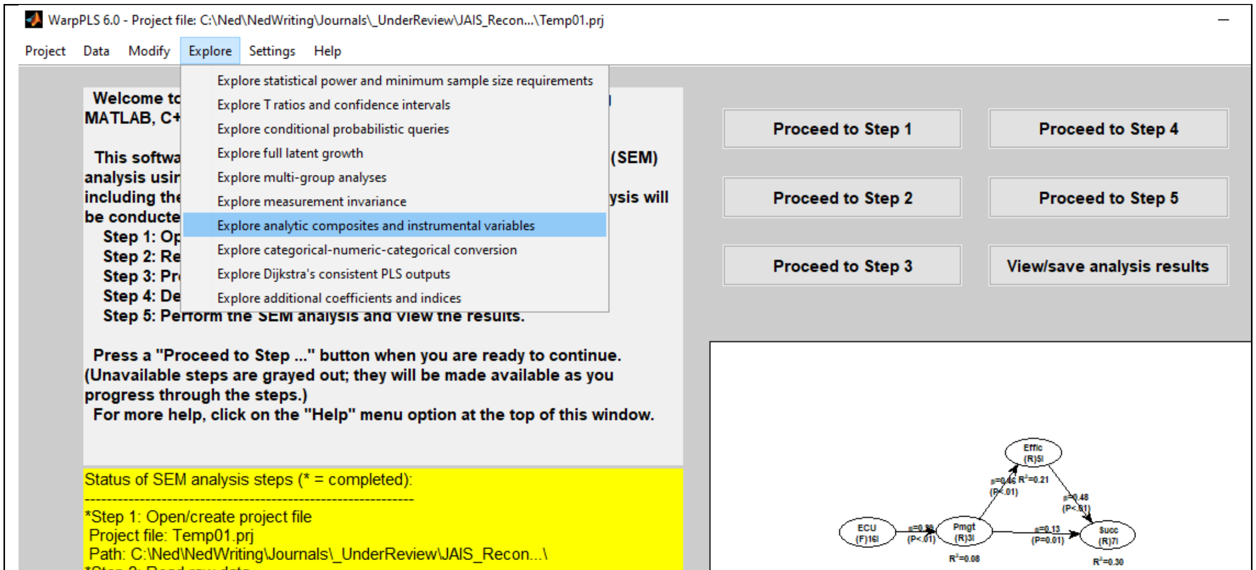

Figure B.1. Analytic composites creation menu option in WarpPLS

Analytic composites are implemented in WarpPLS as weighted aggregations of indicators where the relative weights are set by the user, usually based on one or more existing theories. In WarpPLS relative weight values from -1 to 1 are allowed. For example, an analytic composite may be defined as the aggregation of 3 indicators where the first indicator’s weight is approximately twice that of the second, which is in turn twice that of the third. Here only relative weights matter because the actual standardized weights will be calculated by the software. For instance, assigning the relative weights as 1, 0.5 and 0.25 (each weight being the preceding weight divided by 2) has the same effect as assigning them as 0.3, 0.15 and 0.075.

Figure B.2 illustrates this process. It shows the analytic composite creation screen, which becomes available when the menu option “Explore analytic composites and instrumental variables” is selected. At the top we see a list of all indicators available, of which three are chosen: “ECUEmail”, “ECUEmaillist”, and “ECUBoard”. The following relative weights are assigned to these indicators, respectively: 1, 0.5 and 0.25. The software then aggregates the three indicators according to the relative weights assigned, standardizes the result, and recalculates the weights. These are now the standardized weights: .665, .333, and .166. Once the “Create” button is clicked, an analytic composite is created by the software with the name selected. The default option is “ac_NewName”, which can be changed to a more descriptive name. The analytic composite that is created can then be used in the model for a subsequent analysis; as a single indicator of a latent variable, or as one of the indicators of a multi-indicator latent variable.

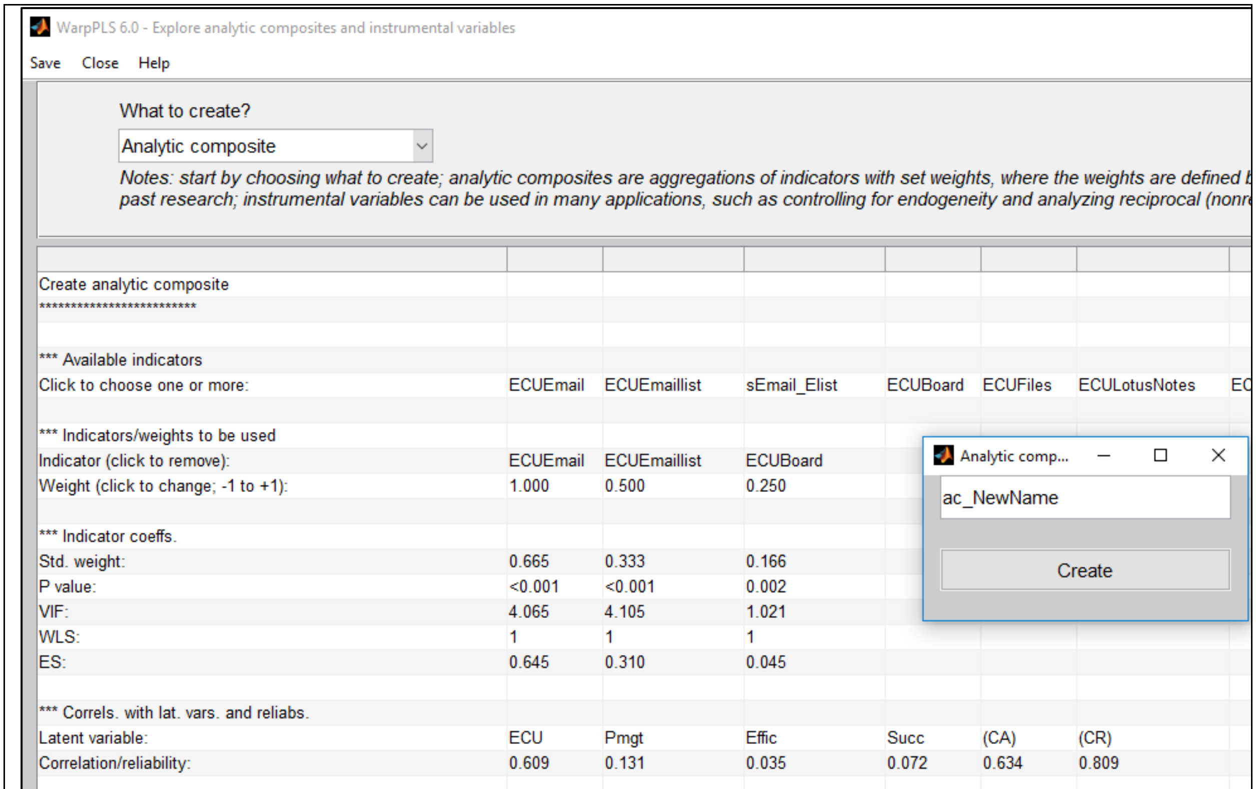

Figure B.2. Selecting analytic composite indicators and weights in WarpPLS

Also shown are the P values for the weights, in the row labeled "P value"; the variance inflation factors for each indicator, in row "VIF"; weight-loading signs, in row "WLS", where a -1 indicates a weight and loading of different signs; and the effect sizes associated with each indicator, in row "ES"; and correlations with each of the latent variables in the model or reliabilities, in row "Correlation/reliability". The reliabilities provided are the Cronbach's alpha (under the label "CA") and the composite reliability (under "CR").

The reliability measures mentioned above are estimates calculated through the classic equations defining the Cronbach's alpha (see, e.g.: Nunnaly,1978; Nunnally & Bernstein, 1994) and the composite reliability (see, e.g.,: Dillon & Goldstein, 1984; Peterson & Yeolib, 2013). These reliability measures should not be confused with the true factor reliabilities discussed earlier in this paper. Strictly speaking, all composites have a reliability of 1. Therefore, these measures should be seen as pseudo-reliabilities.

Appendix C: Indicators and question-statements

A Likert-type scale (0 = “Strongly Disagree” to 10 = “Strongly Agree”) was used for each of the indicators listed below. The latent variable “electronic communication media use (ECM)” was measured formatively, with different indicators assumed to capture different dimensions of the corresponding underlying mental idea, and refers to electronic communication tools used in a team project. All other latent variables were measured reflectively, with different indicators assumed to capture the one main dimension of the corresponding underlying mental idea. The latent variable “team success (Success)” refers to the product developed by the team.

Electronic communication media use (ECM)

ECM1. E-mail to fellow team members (1 to 1).
ECM2. E-mail to team distribution lists (1 to many).
ECM3. Team messaging boards or team discussion forums.
ECM4. Shared electronic files.
ECM5. Share electronic workspace to facilitate sharing information among team members.
ECM6. Electronic newsletters that covered project information.
ECM7. Auto routing of documents for team member and management approval.
ECM8. File transfer protocols (FTP) to attach documents to e-mails and Web pages.
ECM9. A Web page dedicated to this project.
ECM10. A Web page for this project that contained project specs, market research information, and test results.
ECM11. Voice messaging.
ECM12. Teleconferencing.
ECM13. Video conferencing
ECM14. Desktop video conferencing
ECM15. Attaching audio files to electronic documents.
ECM16. Attaching video files to electronic documents.

Project management techniques use (Prjmgt)

Prjmgt1. The team followed a clear plan -- a roadmap with measurable milestones.
Prjmgt2. There were adequate mechanisms to track the project's progress.
Prjmgt3. There were adequate mechanisms to track the project's costs.

Product innovation efficiency (Effic)

Effic1. The product was launched within or under the original budget.
Effic2. The product came in at or below cost estimate for development.
Effic3. The product came in at or below cost estimate for production.
Effic4. The product was launched on or ahead of the original schedule developed at initial project go-ahead.
Effic5. Top management was pleased with the time it took us from specs to full commercialization.

New product success (Success)

Success1. Met or exceeded volume expectations.

Success2. Met or exceeded sales dollar expectations.

Success3. Met or exceeded the 1st year number expected to be produced and commercialized.

Success4. Overall, met or exceeded sales expectations.

Success5. Met or exceeded profit expectations.

Success6. Met or exceeded return on investment (ROI) expectations.

Success7. Met or exceeded overall senior management's expectations.