WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

ePub^{WU} Institutional Repository

Horst Treiblmaier and Peter M. Bentler and Patrick Mair

Formative Constructs Implemented via Common Factors

Article (Accepted for Publication)

Original Citation:
Treiblmaier, Horst and Bentler, Peter M. and Mair, Patrick (2011) Formative Constructs Implemented via Common Factors. *Structural Equation Modeling*, 18 (1). pp. 1-17. ISSN 1070-5511

This version is available at: http://epub.wu.ac.at/3040/
Available in ePub^{WU}: March 2011

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is the version accepted for publication and — in case of peer review — incorporates referee comments. There are minor differences between this and the publisher version which could however affect a citation.

http://epub.wu.ac.at/

Formative Constructs Implemented via Common Factors

Horst Treiblmaier

Vienna University of Economics and Business Administration

Peter M. Bentler

University of California, Los Angeles

Patrick Mair

Vienna University of Economics and Business Administration

ABSTRACT

Recently there has been a renewed interest in formative measurement and its role in properly specified models. Formative measurement models are difficult to identify, and hence to estimate and test. Existing solutions to the identification problem are shown to not adequately represent the formative constructs of interest. We propose a new two-step approach to operationalize a formatively measured construct that allows a closely-matched common factor equivalent to be included in any structural equation model. We provide an artificial example and an original empirical study of privacy to illustrate our approach. Detailed proofs are given in an appendix.

Keywords: Formative Measurement, Identification, Structural Equation Modeling, Latent Variables, Canonical Correlation

INTRODUCTION

Since the time of Spearman (1927) and Thurstone (1947), a common factor has been considered to be an unmeasured variable whose variation (1) generates variation in and (2) explains the correlations or covariances among two or more dependent variables that it predicts (e.g., Mulaik, 2009). The best known examples are first-order factors that explain covariation among observed variables and second-order factors that explain covariation among first-order factors. The part of an observed variable not generated by common factors is a unique factor. From this viewpoint, "Latent variables (LVs) are simply common or unique factors" (Bentler, 1980, p. 423). More recently, proposals have been made to define a different class of LVs that are neither common nor unique factors (e.g., Bollen & Lennox, 1991). Such non-factor LVs are specified to be generated by observed variables rather than the other way around. We will refer to these non-factor LVs as formative LVs. It is well known that formative LVs are hard to identify (e.g., Bagozzi, 2007; Edwards & Bagozzi, 2000; Howell, Breivik, & Wilcox, 2007ab; MacCallum & Browne, 1993) and hence difficult to use in modern multivariate modeling. The aim of this paper is to clarify this difficulty and to propose a new methodology to approximate a non-factor LV by an identified common factor that can be used in any structural equation model (SEM) in the standard way (Bentler, 2010; Iacobucci, 2009, 2010; Savalei & Bentler, 2006).

Let $\mathcal{F}$ represent a formative LV and F represent a common factor LV. We follow MacCallum and Browne (1993) in clearly distinguishing between Fs and $\mathcal{F}$s, which they call "latent" and "composite" variables. We evaluate the extent to which these may be related and provide a new methodology for unambiguously implementing an F that will closely approximate an $\mathcal{F}$. If V is an observed variable, "reflective" measurement models have defining $F \rightarrow V$ paths while "formative" measurement models have defining $V \rightarrow \mathcal{F}$ paths. The Vs in the former case

may be called reflective indicators and in the latter case, formative or causal indicators, and the models, “reflective” and “formative” or “causal indicator” measurement models. Clearly, the defining equations of the two approaches reverse the roles of predictor and criterion variables. An important corollary of a well-defined F is that it generates a set of dependent Vs that are correlated, ideally highly correlated; while a set of Vs need not be correlated at all to define a formative $\mathcal{F}$.

An important distinction that is frequently glossed over in the literature is that a V → $\mathcal{F}$ path is not a V → F path. Although the LISREL model (Jöreskog, 1977) did not explicitly allow V → F paths, these have been common in the SEM field since the Bentler-Weeks (1980) model made explicit provision for such paths. In SEM there is no assumption or requirement that predictor Vs determine the meaning of a consequent F. In contrast, in formative measurement the meaning of the $\mathcal{F}$ is presumed to be determined by the Vs: “the indicators determine the latent variable” (Bollen & Lennox, 1991, p. 306).

Before proceeding with a more detailed summary of approaches to formative measurement and a description of our methodology, we review some relevant issues.

LATENT VARIABLES AND CONSTRUCTS

There is no unanimity in the recent literature on the meaning of “latent variable” and “construct,” and whether, and to what extent, these may have the same meaning. Historically, constructs were equated with common factors, and hence are LVs. According to Cronbach and Meehl (1955, p. 283) “A construct is some postulated attribute of people, assumed to be reflected in test performance.” This viewpoint implies the use of a reflective measurement model, such as that of factor analysis or traditional SEM, where the LVs generate the observed variables. From

this viewpoint, an LV is the common factor that operationalizes a construct. Such LVs must increase the space of variables to larger than the space of observed variables; for p observed variables, there must be at least p+1 linearly independent variables in the model (Bentler, 1982).[1] Conceptually, unmeasured variables that operationalize a construct in a "formative" measurement model also are LVs under appropriate conditions -- that is, they expand the space of variables. However, as we will see, providing an operationalization that does not confound $\mathcal{F}$s with Fs is quite a challenge. In our opinion, such confounding is standard practice in formative measurement.

Not everyone has accepted Cronbach and Meehl's definition of a construct. For example, there is discussion of whether given scientific constructs are "inherently" formative or reflective, that is, whether the substantive meaning of a given construct implies that a researcher has no option but to operationalize it as either an $\mathcal{F}$ or an F. This is a view taken, for example, by Bollen and Lennox (1991), Jarvis, MacKenzie, and Podsakoff (2003), and Podsakoff, MacKenzie, Podsakoff, and Lee (2003). Others propose that "constructs themselves, posited under a realist philosophy of science as existing apart from their measurement, are neither formative nor reflective" (Wilcox, Howell, & Breivik, 2008, p. 2). We lean to the latter viewpoint, but then, there is no need to take a stand on this. Any scientist who has an interest in formative constructs can implement our proposed methodology, regardless of their specific view of the nature of constructs.

[1] Other views of LVs are possible. Bollen (2002) provides an excellent overview.

FACTOR SCORE PREDICTORS AND CONSISTENCY AT LARGE

It is tempting in both measurement models to substitute good estimates $\hat{\mathcal{F}}$ and $\hat{\mathrm{F}}$ for the actual LVs $\mathcal{F}$ and F, because then all subsequent structural regressions of interest can be simply carried out and interpretational confounding can be minimized (Burt, 1976). Skrondal and Laake (2001) provide the most sophisticated approach based on factor score estimates, but it is limited to three groups of factors and allows no higher-order factors. In the context of partial least squares (PLS) estimation of formative measurement models, such substitution of estimates for the true LV is routinely done but with the consequence that "in general, not all parameters will be estimated consistently" (Dijkstra, 2010, p. 37). Of course in both situations, consistent estimates of the LVs could be obtained as the number of observed indicators goes to infinity (a property known as "consistency at large" in the context of PLS), but this theoretical result does not help in practice.

In the following, we do not take a stand on whether formative or reflective measurement models are the "right" way to approach measurement in any given content domain. Consistent with standard SEM practice, we certainly believe that it is appropriate at all times, when it is theoretically meaningful, to predict an F from a set of independent Vs, that is, to have V→F paths in a larger model. The literature has provided no mathematical or statistical criterion to decide when an F so predicted may be interpreted as an $\mathcal{F}$, although substantial thought has gone into the logical and conceptual issues that would help to differentiate these two model types (see, e.g., Coltman, Devinney, Midgley, & Veniak, 2008; Edwards & Bagozzi, 2000; Hardin, Chang, & Fuller, 2008; Howell et al., 2007ab; Jarvis et al., 2003; Kim, Shin & Grover, 2010; MacKenzie, Podsakoff, & Jarvis, 2005; Petter, Straub, & Rai, 2007; Wilcox et al., 2008). For brevity, we assume that the important content and substantive issues have been appropriately

addressed, and the researcher has a good rationale for being interested in $\mathcal{F}$ and understands how to interpret it (e.g., Cenfetelli & Bassellier, 2009). This permits us to focus on the problem of identifying and obtaining $\mathcal{F}$ in practice, to which we now turn.

FORMATIVE MEASUREMENT MODELS

Formative measurement models, by themselves, are not identified. In particular, the unknown coefficients b_i in an equation such as $\mathcal{F} = b_1V_1 + b_2V_2 + \ldots + b_kV_k$ for k predictors of $\mathcal{F}$ cannot be determined, because $\mathcal{F}$ is unobserved. It is possible to define $\mathcal{F}$ =P, a principal component or any other linear combination of variables and to take $P = b_1V_1 + b_2V_2 + \ldots + b_kV_k$, where now b_i is the weight for a variable to generate the component. However, this choice is inadequate, since in this case P is not an LV. It can just be directly computed and put into the data file. This would not be possible with a true LV. For this reason, it is recognized that a formative construct operationalized as an LV also must have a disturbance or residual term.[2] Hence we require that $\mathcal{F} = P + D$, where D is a random disturbance uncorrelated with P. Unfortunately, the random D is unknown, leading to the problem that the formative construct $\mathcal{F} = P + D$ is not uniquely defined. This is the well-known problem of identification. What, then, is a researcher to do to obtain a meaningful $\mathcal{F}$?

[2] In early terminology, "formative" measurement did not require a disturbance term, while the phrase "causal indicator" is meant to permit but not require the disturbance. Issues regarding this disturbance term are summarized in Diamantopoulos (2006) and Wilcox et al. (2008).

PRIOR METHODS TO IDENTIFY FORMATIVE CONSTRUCTS

Solutions to the identification problem proposed by prior writers require expanding the formative measurement model by adding reflective indicators to every formative construct in the model (e.g., Franke, Preacher, & Rigdon, 2008; MacCallum & Browne, 1993; Diamantopoulos &Winklhofer, 2001; Jarvis et al. 2003, p. 214-215; Petter et al. 2007, pp. 640-643). For a practical application, see e.g., Roberts and Thatcher (2009) or Kim, Shin, and Grover (2010). This can be done in several ways, as illustrated in Figure 1. In each case, $\mathcal{F}$ is taken to have reflective indicators (for simplicity, only two in each figure), whether factors (part A), observed variables (part B), or a combination of the two (part C). The figures are simplified to illustrate the main point; see Jarvis et al. (2003, p. 214 Figure 5, Panel 2), Diamantopoulos, Riefler, and Roth (2008), or Bollen and Davis (2009a) for additional details.[3] In each case $\mathcal{F}$ is identified because it has two arrows emanating from it to different dependent variables (Bollen & Davis, 2009ab). It is presumed that these paths are nonzero and the problem of two-indicator identification has been addressed.

The three approaches to identification of $\mathcal{F}$ in Figure 1 are discussed in the literature (see e.g., Jarvis et al., 2003; MacCallum & Browne, 1993; MacKenzie et al., 2005; Petter et al., 2007), while more complicated cases are also considered by Bollen and Davis (2009a). Since $\mathcal{F}$ in each part of the figure is meant to be a formative construct, it also must be predicted by one or more of the Vs that are intended to create $\mathcal{F}$, with a disturbance, say D. That is, we also have $\mathcal{F} = b_1V_1 + b_2V_2 + \ldots + b_kV_k + D$.

[3] Figure 1 appears to show that $\mathcal{F}$ is an independent variable, but it is in fact a dependent variable via $\mathcal{F} = P + D$. The notation with V, F, E, and D is that of EQS (Bentler, 2006).

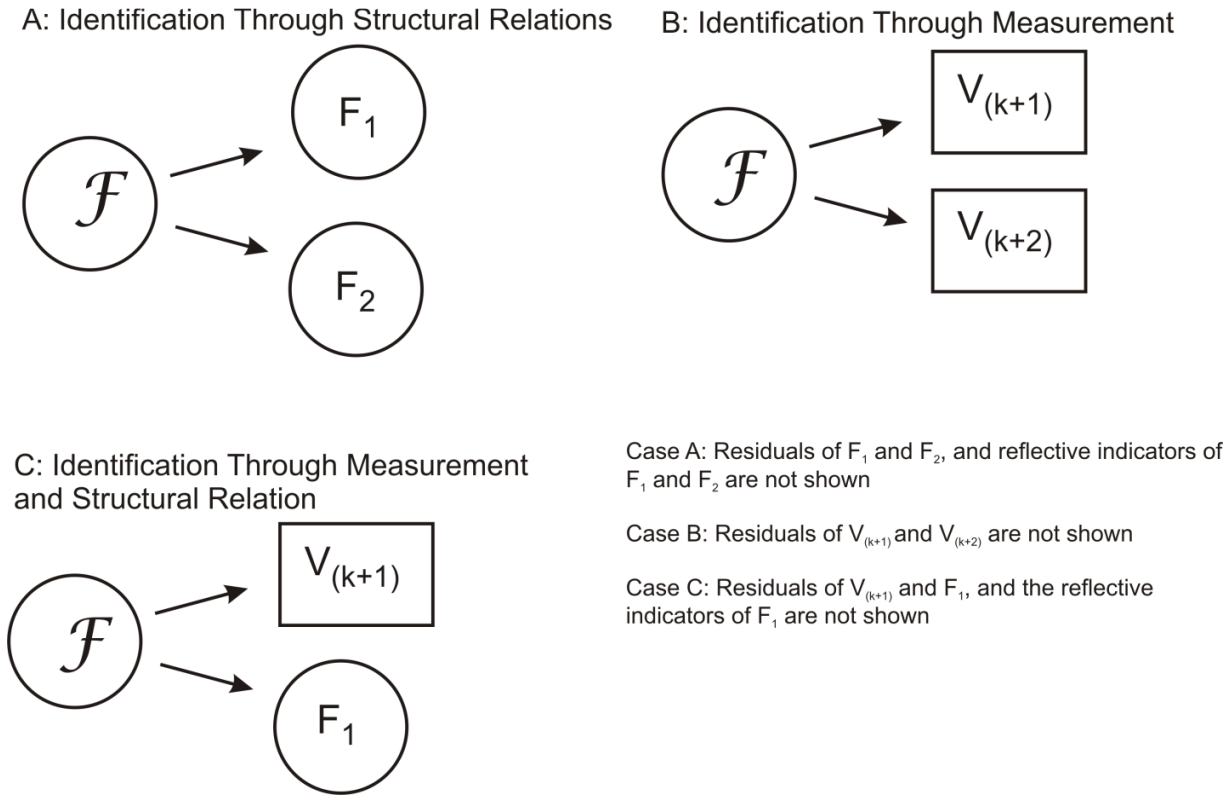

Figure 1. Proposed Solutions for Construct Identification

However, an oddly overlooked point is that *while these proposed solutions provide a model that is identified, they also change the presumed formative* $\mathcal{F}$ *into a reflective* F. To see why this is so, we use the Thurstonian view of a common factor. Consider case A. If F_1 and F_2 are ordinary first-order factors with their own V indicators, then by covariance algebra or path tracing $\mathcal{F}$ is just the second-order factor that explains the correlation between F_1 and F_2. The meaning of this presumed $\mathcal{F}$ is entirely derived from the common variance shared by F_1 and F_2, i.e., the $\mathcal{F}$ in the figure should be replaced by an F. Clearly its meaning is not derived from the fact it is predicted by some Vs, since $\mathcal{F}$ would disappear if F_1 and F_2 were uncorrelated. Yet, the whole point of a formative LV like $\mathcal{F}$ is that its meaning should be determined by its predictors, not by its consequences.

The same problem occurs, and more obviously so, in case B. Here $\mathcal{F}$ is a standard first-order factor F that simply reflects the common variance that V_{k+1} and V_{k+2} share. The fact that there may be $V_1 \ldots V_k$ that predict this standard reflective factor, which we would include in the model when believing $\mathcal{F}$ to be a formative factor, does not matter. Critically, if V_{k+1} and V_{k+2} happen to be uncorrelated, $\mathcal{F}$ will disappear. Again, it makes no sense to consider the existence of a formative LV to depend on the extent to which its consequences correlate, since formative constructs are supposedly created by their observed precursor Vs. If instead of using V_{k+1} and V_{k+2} as indicators in Figure 1b we had used any pair of $V_1 \ldots V_k$, $\mathcal{F}$ also would become a reflective factor in the classical sense.

None of these critical issues is alleviated by using "mixed" indicators as in case C. Rather than being created by its indicators; $\mathcal{F}$ again simply reflects the shared variance between V_{k+1} and F_1. If V_{k+1} and F_1 were to correlate zero, $\mathcal{F}$ would not exist since its reflective paths simply reproduce the correlation of V_{k+1} and F_1.

While our critique is formal, based on the matrix theory behind path tracing, this point has been recognized empirically. Howell et al. (2007a) illustrate to what extent the relation between the formative indicators and a hypothesized construct depends on the indicators of downstream constructs, i.e., those that are supposed to depend on the formative LV. They empirically show that "changing dependent constructs changes the formative construct" (p. 11)[4].

[4] This problem also pertains to reflective measurement, but with less severity. As Howell et al. (2007a, p. 208) state "No, reflective measures are not immune from contamination and potential interpretational confounding when estimated in a larger structural equation model. However, reflective models have epistemic relationships that exist independently from structural relationships." See also Kim et. al. (2010).

Similarly, Kim et al. (2010, p. 353) note that “even when the construct and measures are correctly specified, their relationship changes depending on the choice of endogenous variables.”

The consequences of this unacknowledged transformation of a priori $\mathcal{F}$s into research-usable Fs are known. In their reply to Bagozzi (2007) and Bollen (2007), Howell et al. (2007b, p. 245) summarize that "We do believe that there is one overriding message in this exchange. To paraphrase and extend Bagozzi's observation, for a formatively measured variable, the parameters relating the latent construct to its indicants, as well as the error term in the model, will always be a function of the observed relationships between the formative indicators and the outcome variables, whether interpretational confounding occurs or not." This observed relationship is proportional, as noted by Franke et al. (2008).

Our clear distinction between $\mathcal{F}$s and Fs is theoretical, not empirical, as a thought example can illustrate. Suppose that $\mathcal{F} = b_1V_1 + b_2V_2 + b_3V_3 + b_4V_4 + D$ is the intended formative construct. A meaningful construct in information science is given by Petter et al. (2007, p. 637), namely the construct $\mathcal{F}$ of Declarative Knowledge (from Yi & Davis, 2003), predicted by ability to understand tasks in Excel, such as copying formulas (e.g., V_1), copying contents of cells (V_2), using menu shortcuts (V_3), and using operators (V_4). Now suppose V_5 and V_6 are criterion measures such as capability to set up a new spreadsheet and correctness in analyzing an old spreadsheet. No doubt V_5 and V_6 would be highly correlated, so they would be good indicators of a latent factor F. Furthermore, because of the content similarity of the formative definers V_1-V_4 of $\mathcal{F}$ with the outcomes V_5 and V_6, in a model with F in place of $\mathcal{F}$ we would expect the coefficients $b_1 \ldots b_4$ to be substantial in magnitude. We would congratulate ourselves for having a strong formative construct. But now suppose instead of the effect measures that relate to work with spreadsheets, we let V_5 and V_6 be measures of participants’

physical height and weight. Again, we would expect V_5 and V_6 to be substantially correlated, and they would be good effect indicators of a latent construct F. But surely now the coefficients $b_1 \ldots b_4$ that presumably define F in a formative way would be essentially zero, since it is unlikely that understanding tasks in Excel is related to the height and weight of the respondents. Clearly, F has completely changed meaning, which it should not do if F really were the formative construct $\mathcal{F}$ whose meaning is determined by its antecedents V_1-V_4 rather than by its consequences V_5 and V_6.

Recognizing the difficulties in current approaches to operationalizing an $\mathcal{F}$, next we propose a new method for obtaining a formative LV.

A TWO-STEP APPROACH TO FORMATIVE CONSTRUCT IDENTIFICATION

Consider again the formative construct equation $\mathcal{F} = b_1V_1 + b_2V_2 + \ldots + b_kV_k + D$. The proposed two-step approach to identifying this model involves, in the first step, partitioning P into two or more parts, such as with $P=P_1+ P_2$, and in the second step, defining a latent variable F based on P_1 and P_2 that is as close as possible to $\mathcal{F}$. This is necessary, since $\mathcal{F}$ is inherently unpredictable.

Proposition 1. *$\mathcal{F}$ is linearly predictable from F, with a squared correlation $\rho^2_{\mathcal{F}F}$ that depends on the inherent relation of $\mathcal{F}$ with F and the magnitudes of the disturbance variance in $\mathcal{F}$ and the error variance of its observed component.* (For proof, see Appendix A.)

The P to be split has to be defined. We accept any method that the researcher considers appropriate based on the formative literature. To illustrate, a principal component analysis (PCA)

may define the coefficients, so that $P = b_1V_1 + b_2V_2 + \ldots + b_kV_k$, where P is the component associated with the largest eigenvalue of the predictor correlation or covariance matrix (depending on one's theory). Alternative traditional choices for the weights b_i are those that define the centroid of the predictor variables, or those from a completely a priori weighting scheme. The use of a priori weights based on meta-analysis for identifying optimum weights was recently proposed by Hardin et al. (in press). Unfortunately, even a correctly specified P is incomplete since it provides no rationale for choice of a disturbance term D. We will therefore find a way to proceed without D.

Step 1: Create Multiple Composites

Given that P has been defined, and if the Vs that compose P are not all mutually uncorrelated, we may split P into multiple composites in one of (at least) four different ways: an a priori method, split with original weights, split with replicated weights, and split with weights for maximal correlation. We illustrate with two parts, P_1 and P_2. These methods are simple suggestions for obtaining multiple indicators of a single latent variable, an issue that reflective measurement and LV SEM has dealt with successfully for decades.

<u>A priori split</u>. Since $P = b_1V_1 + b_2V_2 + \ldots + b_kV_k$ we may partition P based on theory and experience into two sets. An example is odd and even sums $P_1 = b_1V_1 + b_3V_3 + \ldots + b_iV_i$ and $P_2 = b_2V_2 + b_4V_4 + \ldots + b_kV_k$. While formative measurement allows the parts P_1 and P_2 to be

uncorrelated, as will be seen below, for our purposes the best split would yield highly correlated parts.[5]

Split with original weights. The original variables defining P are split such that each original $V_i = V_{i1} + V_{i2}$. This creates two sets $V_{11}, V_{21}, \ldots, V_{k1}$ and $V_{12}, V_{22}, \ldots, V_{k2}$. This can be done easily if each V_i is a composite such as a test score. For example, a total test score could be decomposed into two components V_{i1} and V_{i2} by placing half the items into one set, and the other half into the other set by a random choice or by a systematic procedure such as odd vs. even items, and creating subtotal scores based on such a division. Then, using the original weights, $P_1 = b_1V_{11} + b_2V_{21} + \ldots + b_kV_{k1}$ and $P_2 = b_1V_{12} + b_2V_{22} + \ldots + b_kV_{k2}$. Use of identical weights presumably will lead to highly correlated parts.

Split with replicated weights. The chosen weighting procedure (e.g., PCA) can be applied separately to the two sets of variables, yielding $P_1 = b_{11}V_{11} + b_{21}V_{21} + \ldots + b_{k1}V_{k1}$ and $P_2 = b_{12}V_{12} + b_{22}V_{22} + \ldots + b_{k2}V_{k2}$. This case allows instability of the weights to adjust the definition of $P \cong P^* = P_1 + P_2$, and may yield parts that are less highly correlated.

Split with weights for maximal correlation. In this approach, given $V_i = V_{i1} + V_{i2}$ as given above, weights are obtained such that $P_1 = b_{11}V_{11} + b_{21}V_{21} + \ldots + b_{k1}V_{k1}$ and $P_2 = b_{12}V_{12} + b_{22}V_{22} + \ldots + b_{k2}V_{k2}$ are maximally correlated. This is a standard problem in canonical correlation analysis, a well-known technique (e.g., Raykov & Marcoulides, 2008). Again, $P^* = P_1 + P_2 \cong P$.

All of the above methods can be implemented when it is desired to partition P into three or more parts, e.g., with canonical correlation (Kettenring, 1971). When it is literally impossible

[5]A simple example from Bollen (2002, p. 616) may help. He suggests "time spent with friends, time spent with family, and time spent with coworkers as indicators of the [formative] latent variable of time spent in social interaction." If the indicators are V_1, V_2, and V_3, and $P = V_1 + V_2 + V_3$, we could take $P_1 = V_1 + V_2$ and $P_2 = V_3$.

to partition P into any parts, we may use an estimate of the reliability of P to accomplish the same goal as given in Step 2 (see Discussion section).

Step 2: Define F from P_1 and P_2

Next, the composites P_1 and P_2 are treated as reflective indicators of an F; that is, $P_1 = \beta_1 F + E_1$ and $P_2 = \beta_2 F + E_2$. We assume that E_1 is uncorrelated with E_2. Clearly, F is the common part of P_1 and P_2 , i.e., it is a latent variable approximation to P, the determinate part of $\mathcal{F}$. This F can now be incorporated into any larger substantive model. If only two composites P_1 and P_2 are used, standard cautions to identify two-indicator factors must be taken.

Figure 2 summarizes the two steps. Grey shaded areas represent the first step, and the remainder the second step.

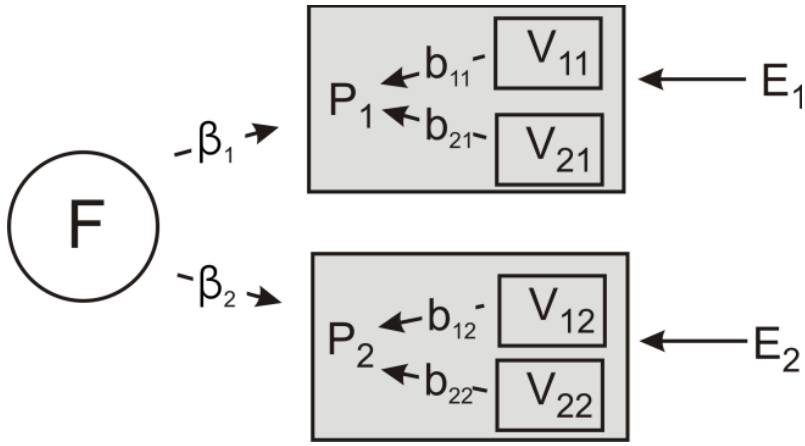

Figure 2. Two-Step Identification with Composites

When there are two P_i we might set $\beta_1=\beta_2$ if the items are split into two equivalent sets. The βs relate the observed composites to a latent reflective F, and the error terms allow the

observed composites to contain a random residual part. While F is the common factor underlying P_1 and P_2, the size of the correlation between F and $\mathcal{F}$ depends on the choice of method in Step1.

Proposition 2. *In Step 2, the squared correlation* $\rho^2_{\mathcal{F}F}$ *between* $\mathcal{F}$ *and F is maximized when* $\gamma = \beta_1 + \beta_2$ *is maximal, and, for a given level of random error variance, the specific variances in* P_1 *and* P_2 *are minimized.* For proof, see appendix.

In other words, F is a better approximation to $\mathcal{F}$ as P_1 and P_2 are highly correlated and the factor loadings β_1 and β_2 are as large as possible. This favors any splitting method that allows P_1 and P_2 to be highly similar in content and statistical properties. In practice, however, we will not know the exact value of $\rho^2_{\mathcal{F}F}$ in any particular application, since the variance of the disturbance D is not estimated in our procedure. Once F is identified, it can be used in any structural model in the usual way.

ILLUSTRATION OF A PRIORI AND CANONICAL FORMATIVE LVS

We illustrate our approach by using a simple example that permits us to detail the computations needed for implementing our method.

Table 1

Correlations among Six Variables Grouped into Two Sets of Three Variables

	V_{11}	V_{21}	V_{31}	V_{12}	V_{22}	V_{32}
V_{11}	1.0					
V_{21}	0	1.0				

V_{31}	0	0	1.0			
V_{12}	.8	0	0	1.0		
V_{22}	0	.6	0	0	1.0	
V_{32}	0	0	.4	0	0	1.0

We use the correlation matrix given in Table 1, which shows the correlations among six variables that have been grouped into two sets of variables, each of which is intended to form an observed composite in step 1 of our method. Formally, the correlation structure in Table 1 can be shown as a block matrix in terms of

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

We force the variables within a set to be uncorrelated, that is, $\Sigma_{11} = I$ and similarly $\Sigma_{22} = I$. This not only simplifies the computations, but also accounts for a major point made in the formative measurement literature, namely, that the variables used to generate formative constructs need not be correlated at all. Uncorrelatedness precludes Σ_{11} or Σ_{22} from defining reflectively measured constructs. Additionally, we suppose that the cross-correlations between pairs of variables V_{i1} and V_{i2} are not zero.

First we consider defining an F to approximate a formative $\mathcal{F}$ using arbitrary a priori weights of .5, .6, and .7 for each of the composites (step one). Notice that if the within-set correlations were nonzero, such weights might be appropriate to generate a principal component or similar weighted sum. Now simple algebra of expectations will allow us to compute the variances and covariances of $P_1 = .5V_{11} + .6V_{21} + .7V_{31}$ and $P_2 = .5V_{12} + .6V_{22} + .7V_{32}$. It can

easily be determined that the variances of P_1 and P_2 are 1.1, and the covariance between P_1 and P_2 is .612. As a result, the correlation between P_1 and P_2 is approximately .5564. Thus if we add P_1 and P_2 to any data file destined for a larger model, the subpart associated with our new F computed using the model of Figure 2 in step 2 obtains factor loadings of $\beta_1 = \beta_2 = \sqrt{.612}$ if we force the factor loadings to be equal, due to equivalent item splits. If we proceed similarly but first standardize P_1 and P_2 to unit variance, the loadings would equal $\beta_1 = \beta_2 = \sqrt{.5564} \approx .746$. This would be a fairly well-defined F. However, we can do even better than that by calculating the canonical correlation, which can be computed at $\rho = .8$. This value is larger than the correlation of composites of .5564 that we obtained earlier. As a result, using our two-step methodology leads to standardized factor loadings of .894 for the F that best approximates the formative construct $\mathcal{F}$. No other weighting scheme can achieve more accurate results.

EXAMPLE: PRIVACY

"The right to be let alone" (Warren & Brandeis, 1890), or privacy, is an issue that concerns consumers, especially in this interconnected age (see e.g., Lee, Im, & Taylor, 2008). Various reflective scales have been developed to measure privacy. For example, Smith, Milberg, and Burke (1996) focused on individuals' concerns about organizational practices and developed a 15-item instrument with four subscales (collection, error, secondary use, improper access), and Malhotra, Sung, and Agarwal (2004) developed a scale with three first order dimensions (collection, control and awareness) to measure Internet users' information privacy concerns. We use formative indicators (see Appendix B for rationale) intended to assess various aspects of privacy related to perceived disadvantages of personalized communication.

Content validity was insured by consulting experts for gathering the first set of indicators measuring potential disadvantages of individualized communication. Qualitative Interviews were conducted in which experts (e.g. consumer advocates, consultants and producers of CRM software) talked about their perspectives on individualization. We taped, transcribed and paraphrased passages that contained relevant statements and generalized the paraphrases. Finally, these new statements were combined into an abstraction system and checked against the original transcripts. Following the principles of qualitative content analysis, the goal was to identify as many different opinions as possible rather than to ensure the representativeness of individual statements. The items (see Table 2) were used in an online survey with a convenience sample of 405 Internet users. In order to assess their level of agreement or disagreement with the statements, the respondents used a slider bar ranging from 1 to 100, thus creating metric scales.

Table 2

Items and Response Format

Perceived Disadvantages of Personalized Communication		
Item #	Wording	Measurement/Scale Level
V11	On the Internet, data about me is permanently collected, which I cannot control.	Slider Bar/Metric (1-100) Do not Agree – Fully Agree
V21	I am poorly informed about the use of my data.	Slider Bar/Metric (1-100) Do not Agree – Fully Agree
V12	If I divulge personal data, I lose control over how companies use my data.	Slider Bar/Metric (1-100) Do not Agree – Fully Agree
V22	Personalization leads to an increase in unsolicited advertising messages, since companies know what I am interested in.	Slider Bar/Metric (1-100) Do not Agree – Fully Agree

The first step of our two-step approach is performed in R using the package CCA (Gonzales et al., 2008) for canonical correlation analysis. Appendix C gives a few syntax lines on how CCA can be performed in R and SPSS. The factor analytical second step is computed in EQS. The interrelation of EQS and R can be simplified with the REQS package (Mair, Wu, & Bentler, 2010).

For CCA we specify V_{11} and V_{21} as the first set and V_{12} and V_{22} as the second set (see Table 3). The canonical correlation coefficients between P_1 and P_2 are .464 and .084; only the first correlation coefficient is of interest. The corresponding canonical coefficients (weights) are .0197 (V_{11}) and .0220 (V_{21}) for the first set, and .0292 (V_{12}) and .0176 (V_{22}) for the second set. In CCA, the canonical coefficients are typically not used; the correlations between the variables and the new canonical variates are interpreted instead. These are commonly referred to as loadings and can be interpreted as ordinary factor or component loadings (e.g., Raykov & Marcoulides, 2008). The respective values are .786 (V_{11}), .800 (V_{21}) for P_1 and .851 (V_{12}), .703 (V_{22}) for P_2. By means of the canonical coefficients, the scores P_1 and P_2 can be computed and the first step of our approach is completed. Based on these linear composite scores, we compute the reflective part of the model in the second step. Since a CFA with only two indicators is not identified, we restrict the factor loadings to be equal and get standardized loadings of .681. Subsequently, after calculating the loadings, the privacy construct is now ready to be placed into any larger nomological network in order to test the relationships postulated by theory.

Table 3

Results of CCA and Formative Loadings

		Canonical Correlation	Canonical Coefficients	Loadings	Standardized Loadings

		Coefficient			
P_1	V_{11}	.464	.0197	.786	.681
	V_{21}		.0220	.800	
P_2	V_{12}	.464	.0292	.851	.681
	V_{22}		.0176	.703	

DISCUSSION

Structural equation models routinely use ordinary common factors as latent variables, but it has been argued that non-common factor formative LVs would be more appropriate in many contexts. For example, Jarvis et al. (2003) illustrate that most empirical papers in leading marketing journals (Journal of Consumer Research, Journal of Marketing Research, Journal of Marketing, Marketing Science) used common factors when formative LVs might have been a more appropriate choice. Unfortunately, as has long been recognized, this cannot easily be done due to the fundamental problem of identification (MacCallum & Browne, 1993). The only known solution has been to expand the formative measurement model by adding reflective indicators. An unforeseen consequence is that the formative $\mathcal{F}$ is changed into a standard common factor F that accounts for the covariances of its consequent indicators, i.e., it destroys the intended meaning of the formative LV as one whose meaning derives from its antecedents. It simply transforms a model containing formative LVs into a Bentler-Weeks type SEM model that contains V $\rightarrow$ F paths. For example, MacCallum and Browne's (1993, p. 540) identified formative model should contain an $\mathcal{F}$, but instead it contains an F that simply is a second-order factor that is predicted by some observed variables. If the first-order factors were to be

uncorrelated, the second-order factor and hence the presumed formative LV would disappear. This is contrary to the idea that formative LVs are created by their antecedent Vs.

The solution that we propose for this quandary is first, to acknowledge this problem, and then, to devise a way around it. We propose to substitute the hypothesized $\mathcal{F}$ with an F that is as close to it as possible in a well-defined sense. This idea is implemented with a two-step approach that splits the determinate part of the formative construct into two or more composites and then models these composites as a common factor that can be placed into any desired larger SEM model. To illustrate how this can be done in practice, we report an artificial example and an original empirical study. Canonical correlation analysis is used to calculate the composites.

Step I of our method amounts to creating multiple indicators of a common factor. If there is absolutely no way to create two or more indicators in the first place, or if any two composites that might be created correlate approximately zero, this method cannot be used. Of course, the situation is not hopeless. When all Vs that define P are mutually uncorrelated, we may simply use P by itself along with an estimate of its reliability (obtainable, e.g., by a test-retest method). With a reliability coefficient, the variance of P can be partitioned into that of factor and error. Such a procedure to handle single indicators has been used in latent variable structural modeling for decades and is still considered to be a viable option in recent literature (e.g., Coffman & MacCallum, 2005, p. 240; Kline, 2005, pp. 229-31; MacKinnon, 2008, p. 189).

The proposed methodology, based on linear structural models, can be easily extended to allow for nonlinear measurement models such as nonlinear canonical correlation analysis for ordinal variables. More generally, our methodology can be extended to allow general nonlinear functions such as $P_1 = f(\mathbf{X})$ and $P_2 = f(\mathbf{Z})$ in the first step, followed by a standard factor model in the second step.

Although technical properties of our approach are described in the two propositions and their proofs, given in Appendix A, further mathematical and statistical results remain to be developed. For example, research will have to determine the actual degree of correspondence between $\mathcal{F}$ and F that is achievable in practice, perhaps based on several alternative splitting methods applied under a variety of simulated conditions. The propositions are based on standard classical test theory and factor analytic assumptions (Bentler, 2009), and hence the effect of violating these assumptions is another topic for further research. It also would be useful to determine the extent to which the factors used in traditional methods for formative identification, such as given in Figure 1, agree with the factors obtained in our new approach.

Our approach allows – but certainly does not require -- the formative indicators to contain errors of measurement. Our propositions hold for perfectly reliable formative indicators as well as those that may be error-prone. It is well known that the standard formative measurement model "does not include error terms for the causal indicators" (MacCallum & Browne, 1993, p. 534). This does not imply that all Vs that generate $\mathcal{F}$ in fact contain no measurement error. Such error is just not modeled, although it could be, as shown in Edwards and Bagozzi (2000, Fig. 5B) and Diamantopoulos et al. (2008, Fig. 3). As noted by Kline (2005, p. 229) "scores from a single indicator are quite unlikely to have no measurement error."

Our equations are based on population quantities. In practice, any method is applied in samples, as it was in our example. Step 1 of our method is simply another data preparation step, similar to those routinely used for preprocessing and data reduction, while Step 2 is a standard structural modeling analysis. Hence, currently available statistical theory (e.g., Yuan & Bentler, 2007) can be used to evaluate the statistical properties of our estimates and tests.

This paper proposes a compromise in the formative/reflective controversy. We accept that, conceptually speaking, a formative construct can be scientifically meaningful, but practically and operationally, a thoughtfully developed reflective measurement approach is the most appropriate way to implement it.

ACKNOWLEDGEMENTS

This research was supported by grants 5K05DA000017-32 and 5P01DA001070-35 from the National Institute on Drug Abuse to Peter M. Bentler, who also acknowledges a financial interest in EQS and its distributor, Multivariate Software. This research was also supported by a Hochschuljubläumsstiftung grant from the city of Vienna.

REFERENCES

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

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, *31*, 419-456.

Bentler, P. M. (1982). Linear systems with multiple levels and types of latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp.101-130). Amsterdam: North- Holland.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74* (1), 137-143.

Bentler, P. M. (2010). SEM with simplicity and accuracy. *Journal of Consumer Psychology*, *20*, 215-220.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289-308.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53* (1), 605-634.

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

Bollen, K. A., & Davis, W. R. (2009a). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling*, *16* (3), 498-522.

Bollen, K. A., & Davis, W. R. (2009b). Two rules of identification for structural equation modeling. *Structural Equation Modeling*, *16* (3), 523-536.

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

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, *5*, 3-52.

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

Coffman, D. L. & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research*, *40* (2), 235-259.

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Diamantopoulos, A. (2006). The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modelling in Management*, 1 (1), 7-17.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*, 1203-1218.

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

Dijkstra, T. K. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang, (Eds.), *Handbook of partial least squares: Concepts, methods, and applications, computational statistics, vol II* (pp. 23-46). Heidelberg: Springer.

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

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research*, *61* (12), 1229-1237.

Gonzales, I., Dejean, S., Martin, P. G. B., & Baccini, A. (2008). CCA: An R Package to Extend Canonical Correlation Analysis. *Journal of Statistical Software*, 23 (12), 1-14.

Hardin, A. M., Chang, J. C., & Fuller, M. A. (2008). Formative vs. reflective measurement: Comment on Marakas, Johnson, and Clay (2007), *Journal of the Association for Information Systems, 9* (9), 519-534.

Hardin, A. M., Chang, J. C., Fuller, M. A., & Torkzadeh, G. (in press). Formative measurement and academic research: In search of measurement theory. *Educational and Psychological Measurement.*

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.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask. *Journal of Consumer Psychology*, *19* (4), 673-680.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20* (1), 90-98.

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

Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P. R. Krishnaiah (Ed.), *Applications of Statistics* (pp. 265-287). Amsterdam: North-Holland.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, *58* (3), 433-451.

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. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Lee, D. –H., Im, S., & Taylor, C. R. (2008). Voluntary self-disclosure of information on the Internet: A multimethod study of the motivations and consequences of disclosing information on blogs. *Psychology and Marketing*, *25*, 692 – 710.

MacCallum, R. C. & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, *114* (3), 533-541.

MacKenzie, S. B., Podsakoff, P. M. & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90* (4), 710-730.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Taylor & Francis.

Mair, P., Wu, E., & Bentler, P. M. (2010). EQS goes R: Simulations for SEM using the package REQS. *Structural Equation Modeling*, *17* (2), 333-349.

Malhotra, N. K., Sung S. K., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale and a causal model. *Information Systems Researc*h, *15* (4), 336-355.

Mulaik, S. A. (2009). *Foundations of factor analysis* (2nd ed.). New York: CRC Press.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31* (4), 623-656.

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

Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis.* New York: Routledge.

Roberts, N. & Thatcher, J. B. (2009). Conceptualizing and testing formative constructs: Tutorial and annotated example. *The DATA BASE for Advances in Information Systems*, *40* (3), 9-39.

Savalei, V., & Bentler, P. M. (2006). Structural equation modeling. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances* (pp. 330-364). Thousand Oaks, CA: Sage.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563-576.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly*, *20* (2), 167-196.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Thurstone, L.. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Warren, S., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, *4* (5), 193-220.

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

Yi, M. Y., & Davis, F. D. (2003). Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research*, *14* (2), 146-169.

Yuan, K.–H. & Bentler, P. M. (2007). Structural equation modeling. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Volume 26: Psychometrics* (pp. 297-358). Amsterdam: North Holland.

APPENDIX A

Let $\mathcal{F}$ be a formatively measured construct and F be a latent factor or any other variable.

Proposition 1. $\mathcal{F}$ is linearly predictable from F, with a squared correlation $\rho^2_{\mathcal{F}F}$ that depends on the inherent relation of $\mathcal{F}$ with F and the magnitudes of the disturbance variance in $\mathcal{F}$ and the error variance of its observed component.

Proof. Let the formatively measured construct be $\mathcal{F}$ = P+D, with P=b_1V_1+b_2V_2+…+ b_pV_p and where V_1 … V_p are observed variables and D is a random disturbance uncorrelated with P. The extent of linear predictability of $\mathcal{F}$ is given by $\mathcal{F} = \gamma F + \varepsilon$ where γ is a coefficient describing the inherent relation of $\mathcal{F}$ with F and ε is a residual uncorrelated with F. The squared correlation between $\mathcal{F}$ and F is $\rho^2_{\mathcal{F}F} = \gamma^2\sigma^2_F / \sigma^2_{\mathcal{F}}$. To allow that P may not be perfectly reliable, we assume that it has a classical test theory decomposition P=T+E, where T (true component) and E (random error, if any) are uncorrelated. Then $\sigma^2_P = \sigma^2_T + \sigma^2_E$, with the possibility that $\sigma^2_E = 0$. It is natural to assume that E is uncorrelated with D, so that $\sigma^2_{\mathcal{F}} = \sigma^2_T + \sigma^2_E + \sigma^2_D$. Hence $\rho^2_{\mathcal{F}F} = \gamma^2\sigma^2_F / (\sigma^2_T + \sigma^2_E + \sigma^2_D)$, and the squared correlation is increased as the variances of E and D decrease. Under formative measurement, $\sigma^2_D > 0$, and maximal predictability is reached when $\sigma^2_E = 0$. Otherwise, maximal predictability occurs when both the disturbance variance σ^2_D and error variance σ^2_E vanish.

Proposition 2. In Step 2, the squared correlation $\rho^2_{\mathcal{F}F}$ between $\mathcal{F}$ and F is maximized when $\gamma = \beta_1 + \beta_2$ is maximal, and, for a given level of random error variance, the specific variances in P_1 and P_2 are minimized.

Proof. Since $P_1 = \beta_1 F + E_1$ and $P_2 = \beta_2 F + E_2$, $\mathfrak{F} = P_1 + P_2 + D = \gamma F + \varepsilon$ where $\gamma = \beta_1 + \beta_2$ and $\varepsilon =$ $(E_1 + E_2 + D)$. The squared correlation is $\rho^2_{\mathfrak{F}F} = \gamma^2 \sigma^2_F / (\sigma^2_F + \sigma^2_{E_1} + \sigma^2_{E_2} + \sigma^2_D)$. For a given denominator, this is maximal when $\gamma = \beta_1 + \beta_2$ is maximal. Also, according to standard factor analytic theory (see Bentler, 2009), the unique variances have decompositions $\sigma^2_{E_1} = \sigma^2_{S_1} + \sigma^2_{\delta_1}$ and $\sigma^2_{E_2} = \sigma^2_{S_2} + \sigma^2_{\delta_2}$ where $\sigma^2_{S_1}$ and $\sigma^2_{S_2}$ represent reliable specific variance not shared between P_1 and P_2, and $\sigma^2_{\delta_1}$ and $\sigma^2_{\delta_2}$ are random error variances. When γ, $\sigma^2_{\delta_1}$ and $\sigma^2_{\delta_2}$ are fixed, $\rho^2_{\mathfrak{F}F}$ is maximized when the specific variances $(\sigma^2_{S_1} + \sigma^2_{S_2}) \to 0$.

APPENDIX B

Following the recommendations from Jarvis et al. (2003), we provide a rationale as to why the items can be classified as being indicators of a formative construct.

Perceived Disadvantages of Personalized Communication	
Decision Rules for Construct Identification	
Decision Rule	Rationale
Direction of causality is from items to construct	The indicators of perceived disadvantages of personalized communication are defining characteristics of the construct; i.e. the construct is made up of the sum total of those cases where an individual perceives a disadvantage. If one indicator changes (e.g. better information about data usage), the whole construct changes.
Indicators need not to be interchangeable	The indicators capture different dimensions of disadvantages, i.e. the loss of control of data collection, the lack of information about data usage, the loss of control of data usage and the increase of unsolicited advertising.
Not necessary for indicators to covary with each other	There are various situations conceivable, where one indicator changes and leaves the rest unchanged. A spam filter might lead to a reduction in unsolicited advertising messages, while leaving the other disadvantages unchanged.
Nomological net for the indicators may differ	The antecedents and consequences for the indicators vary. Legislation, for example, might force companies to inform customers better about data usage practices without changing their marketing strategies. Unsolicited advertising may lead to "information overflow", while poor information about data usage might reduce trust.

APPENDIX C

The CCA package in R written by Gonzales et al. (2008) provides a flexible tool for performing canonical correlation analysis and various extensions. A prototype R syntax could look as follows (comments are quoted after the "#"-sign):

```
install.packages("CCA")                     #install package
library("CCA")                              #load package
set1 <- cbind(item1, item3)                 #set 1: items 1 and 3
set2 <- cbind(item2, item4)                 #set 2: items 2 and 4
res.cca <- cc(set1, set2)                   #CCA
#vector with canonical coefficients
coefvec <- c(res.cca$xcoef[,1],resc.cca$ycoef[,1])

#vector with loadings
loadvec <- c(res.cca$scores$corr.X.xscores[,1],
             res.cca$scores$corr.Y.yscores[,1])

#extract scores
xscores <- res.cca$scores$xscores[,1]       #first score vector
yscores <- res.cca$scores$yscores[,1]       #second score vector
score.mat <- cbind(xscores, yscores)        #produce score matrix

# write score matrix on hard disk.
write.table(score.mat, file = "scoresEQS.dat", row.names = FALSE,
col.names = FALSE)
```

The file "scoresEQS.dat" can be imported directly in EQS for the factor analytical computations. Note that "set1" and "set2" are the responses for the first and second set of items and consequently stored as matrices.

Unfortunately SPSS (PASW) in its current version, does not provide a possibility for CCA computation in the graphical user interface. However, using the "manova" command with the "discrim" option, canonical correlation analysis can be performed.

```
manova item1 item3 with item2 item4
/ discrim all alpha(1)
/ print=sig(eigen dim) .
```

Again, the item 1 and 3 form the first set, item 2 and 4 the second set.