JOURNAL OF Econometrics

Journal of Econometrics 89 (1999) 393–421

Representation of measurement error in marketing variables: Review of approaches and extension to three-facet designs

Richard P. Bagozzi[a,]*, Youjae Yi[b], Kent D. Nassen[c]

[a] *University of Michigan Business School, Ann Arbor, MI 48109-1234, USA*
[b] *Seoul National University, USA*
[c] *University of Michigan, USA*

Abstract

This paper explores approaches for modeling measurement error in marketing research, including random, method and measure specific sources of error. The following approaches are considered: classic confirmatory factor analysis, second-order models, panel models, additive trait-method models, correlated uniqueness models, covariance components analysis, additive trait-method-measure specific-error models, and the direct product model, where traits and methods interact. Finally, a three-facet multiplicative model is addressed wherein latent variables underlying a phenomenon under investigation are shown to interact with multiple methods and occasions of measurement. The three-facet model is illustrated on a study of consumer attitudes toward losing weight explicitly conducted for this paper. © 1999 Elsevier Science S.A. All rights reserved.

Keywords: Construct validity; Multitrait-multimethod matrices; Method error; Confirmatory factor analysis

Questionnaire items and rating scales are used throughout marketing research to operationalize theoretical constructs and make policy decisions. Among other areas, items and rating scales form the basis for measuring perceptions of service quality (Parasuraman et al., 1994), customer satisfaction

* Corresponding author. E-mail:bagozzi@umich.edu

PII: S0304-4076(98)00068-2

(Yi, 1990), evaluations and perceived value of products (Dodds et al., 1991), consumer innovation (Foxall and Haskins, 1986), consumer involvement (Zaichkowsky, 1985), attitudes and beliefs (Bagozzi, 1996), various personality traits or states (Bearden and Rose, 1990), and human judgment and choice (Johnson, 1988).

Any measure based on perceptions, judgments, beliefs, attitudes, or other subjective states is likely to reflect measurement error, as well as the theoretical content or objective information presumed to underlie the psychological reactions or true characteristics of an object under scrutiny. In turn, measurement error can be conceived to consist of random and systematic components.

A common source of systematic error in marketing research is method error. Method error refers to variance attributed to measurement procedures(s) rather than to a construct of interest, and examples include halo effects, social desirability distortions, acquiescence tendencies, evaluation apprehension, demand artifacts, and key informant biases associated with peer or expert ratings (e.g., Bagozzi, 1994a, pp. 26–33; Ganster et al., 1983; Nichols et al., 1982; Phillips and Bagozzi, 1986; Rosenthal and Rosnow, 1969; Seidler, 1974; Winkler et al., 1982).

Econometricians need to be concerned with measurement error because such error can have serious confounding influences on the interpretation of empirical research (e.g., Bagozzi, 1994b, pp. 364–369; Campbell and Fiske, 1959; Fiske, 1982). It is well known that random error frequently attenuates the observed relationships among variables in statistical analyses and therefore may produce errors in inference. Less well known is the possibility that random error can actually inflate parameter estimates under some circumstances in multivariate analyses (e.g., in simultaneous equation systems), depending on the pattern and magnitude of such errors among predictors (Bagozzi, 1994b, p. 365). Likewise, method error may suppress or magnify relationships among variables and contribute to Type I or Type II errors if not taken into account (e.g., Bagozzi et al., 1991). As developed below, we can express four distinct sources of variance in a measure y as follows: var $y =$ var $T +$ var $M +$ var $S +$ var R, where var $T =$ theoretical or 'trait' variance, var $M =$ method variance, var $S =$ specific variance (i.e., systematic variance in y, over and above that attributable to a trait and method), and var $R =$ random error variance.

Because measurement errors (i.e., random error and method variance) provide potential threats to the interpretation of research findings, it is important to validate measures and disentangle influences of these errors in the course of testing hypotheses. This can be achieved by using multiple measures and multiple methods in measurement and hypothesis testing. Using a single measure per variable in a theory under test does not permit one to take reliability into account in analyses. Similarly, with only a single method, one cannot distinguish trait variance from unwanted method variance, because each attempt to measure a concept is potentially contaminated by irrelevant aspects of the method employed.

The purpose of this article is twofold. First, we review a number of procedures that can be used to model measurement error. These procedures consist of simple methods for partitioning variance into trait and error components, as well as more complex methods addressing method and measure specific variation in addition to trait and random error variance. The procedures speak to the issue known as construct validity, which is defined broadly as the extent to which an operationalization measures the concept it is supposed to measure (e.g., Cook and Campbell, 1979). Second, we consider an extension of the direct product model (e.g., Browne, 1984, 1989) to three-facet designs and illustrate the procedure on a set of longitudinal data collected for this purpose. The particular three-facet design we examine models the complete crossing of traits, methods, and occasions of measurement.

1. Modeling measurement error

1.1. Confirmatory factor analysis

Jöreskog (1969) proposed the following 'measurement model' as a representation of the relationship between measures and factors (where certain restrictions must be placed on Λ_y and Θ_ε below to achieve identification):

$$\boldsymbol{y} = \Lambda_y \eta + \varepsilon \tag{1}$$

where $\mathbf{y}$ is a $\boldsymbol{p} \times 1$ vector of observed variables, $\boldsymbol{\eta}$ is an $\boldsymbol{n} \times 1$ vector of latent (i.e., unobserved) variables or factors underlying the observed variables, $\boldsymbol{\varepsilon}$ is a $\boldsymbol{p} \times 1$ vector of error variables, and it is assumed that the $\boldsymbol{\eta}$'s and $\boldsymbol{\varepsilon}$'s are random variables with zero means, $\mathbf{y}$ is measured in deviations from its means, and the $\boldsymbol{\varepsilon}$'s are uncorrelated with the $\boldsymbol{\eta}$'s. The expression Λ_y is a $\boldsymbol{p} \times \boldsymbol{n}$ matrix of partial regression coefficients for the regression of $\mathbf{y}$ on $\boldsymbol{\eta}$ and are commonly referred to as factor loadings. Given the above specification, the implied covariance matrix of $\mathbf{y}$ is

$$\Sigma = \Lambda_y \Psi \Lambda_y' + \Theta_\varepsilon \tag{2}$$

where Ψ is the covariance matrix of $\boldsymbol{\eta}$ and Θ_ε is the covariance matrix for $\boldsymbol{\varepsilon}$ (which for the moment, we will assume is diagonal).

Eqs. (1) and (2) together define the confirmatory factor analysis (CFA) model. For a given specification, and given identification of the parameters in Λ_y, Ψ, and Θ_ε, maximum likelihood procedures have been developed for estimation and various goodness-of-fit measures and related diagnostics have been provided for evaluating models (e.g., Bentler, 1990; Jöreskog and Sörbom, 1993). Leading programs such as AMOS (Arbuckle, 1995), EQS (Bentler and Wu, 1993), and LISREL (Jöreskog and Sörbom, 1993) implement generalized least squares, unweighted least squares, two-stage least squares, instrumental variables, and weighted least squares, including asymptotic distribution free procedures, as well as maximum likelihood estimation.

The CFA model provides a null hypothesis for testing whether covariation amongst a set of measures can be explained by one or more latent variables plus error. Various rules and procedures have been derived for ascertaining identification of such models (e.g., Bollen, 1989, pp. 238–254). In the typical application, either a single latent variable is fit to a set of measures to test for unidimensionality of the measures or multiple latent variables are fit to a set of measures to test for unidimensionality of measures within latent variables and discriminant validity of measures across latent variables (where we assume that each measure is hypothesized to relate to one latent variable only and the assumptions noted above for Eq. (1) are met). Identification under the former case (i.e., the single factor model) will be met when three or more measures are specified (with three measures, the model will be exactly identified; with four or more it will be overidentified). Identification under the latter requires only two measures per latent variable when at least two latent variables are under investigation. More complex models, where measures relate to two or more latent variables and/or covariances among error terms are permitted, can be estimated under certain specific conditions for identification. But with respect to establishing reliability and validity of measures, such models are less interesting (except as noted below) because they introduce ambiguities in interpretation and usually imply specification errors (e.g., omitted variables or alternative dimensionality than hypothesized).

A number of criteria are used to assess goodness-of-fit of CFA models. For maximum likelihood estimation procedures, the chi-square measure can be used as a test statistic for testing a hypothesized model as a null hypothesis. The degrees of freedom for the chi-square test is $df = 1/2(p)(p+1) - t$, where again p is the number of observed variables, $1/2p(p+1)$ is the number of covariances, and t is the number of parameters to be estimated. A nonsignificant chi-square value with associated degrees of freedom is taken as indicating a satisfactory model.

Another approach to the assessment of goodness-of-fit is to use an index which is based on the comparison of the fit of a hypothesized model to the fit of a baseline model, such as a null model in which all variables are uncorrelated (i.e., only error variances are estimated). Such comparisons can be used to calculate incremental fit indices, which contrast the fit of a hypothesized model to that of a more restricted, nested model. Currently, the most widely recommended fit index is the comparative fit index (e.g., Bentler, 1990):

$$CFI = \frac{(\chi_0^2 - df_0) - (\chi_f^2 - df_f)}{\chi_0^2 - df_0} \tag{3}$$

where χ_0^2 and χ_f^2 are for the null and focal models, respectively, and df stands for degrees of freedom. The CFI is normed in the population and thus has values bounded by 0 and 1, except when $\chi_f^2 < df_f$, in which case the convention is to report $CFI = 1.00$ (Bentler, 1990). The CFI provides an unbiased estimate of its corresponding population value, and therefore should be independent of sample size. Monte Carlo studies show that the CFI performed well for sample sizes varying from 50 to 1600, in the sense of producing estimates that were unbiased and low in variability (e.g., Bentler, 1990). The CFI should be interpreted as a measure of how much variation in measures is accounted for from a practical standpoint. A rough rule of thumb is that values of the $CFI \geqslant 0.90$ suggest that further relaxation of parameter constraints is not warranted and might lead to overfitting. However, one should also consider model complexity in interpreting fit indices. For example, relatively low CFI values might occur due to high model complexity (Bone et al., 1989).

A third goodness-of-fit measure that is sometimes insightful is the root-mean-square error of approximation ($RMSEA$). The $RMSEA$ indicates model discrepancy per degree of freedom and thus compensates (i.e., incorporates a penalty) for model complexity. Other goodness-of-fit tests based on discrepancy functions tend to favor models with many parameters. Browne and Cudeck (1993) indicate that values of 0.08 or less for the $RMSEA$ provide evidence for reasonable fits and that values less than or equal to about 0.10 may be satisfactory for exploratory research. Alternatively, we may use the Tucker and Lewis (1973) index, which expresses noncentrality per degrees of freedom: $TLI = 1 - (\chi_f^2 - df_f)(df_0)/(\chi_o^2 - df_0)(df_f)$. The TLI considers the relative parsimony of alternative models and is hardly affected by sample size (Steenkamp and Van Trijp, 1991). However, the TLI is not restricted to fall between 0 and 1 and its sampling variability is larger than that of the CFI (but see Marsh et al., 1996). Bagozzi and Edwards (1998) provide illustrations of the CFA model.

1.2. Second-order CFA models

The CFA model defined by Eqs. (1) and (2) has proved useful in establishing the dimensionality of constructs and scales and in demonstrating the degree of reliability of measures. However, one drawback with the CFA model is that the variance attributed to a trait may contain variance shared by all measures plus variance unique to measures of a specific trait. The first is termed common variance and the second is called specific variance (not to be confused with the notion of measure specificity discussed later). The additional variance associated with common variance will contribute to the reliability of a measure but of course may be misleading, as the researcher is typically most interested in the variance arising from a particular latent variable.

One way to partition the variance in measures into common, random error, and specific variance components is to perform a second-order confirmatory factor analysis (SCFA). The general SCFA model can be expressed through two systems of equations as follows:

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{4}$$

$$\boldsymbol{\eta} = \boldsymbol{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{5}$$

where for the 'measurement model' in Eq. (4), **y** is a $\boldsymbol{p} \times 1$ vector of observed variables, $\boldsymbol{\eta}$ is $\boldsymbol{m} \times 1$ vector of latent variables underlying the observed variables, Λ_y is a $\boldsymbol{p} \times \boldsymbol{m}$ matrix of partial regression coefficients for the regression of **y** on $\boldsymbol{\eta}$, ε is a $\boldsymbol{p} \times 1$ vector of random error variables; and for the 'structural model' in Eq. (5), $\boldsymbol{\xi}$ is a $\boldsymbol{q} \times 1$ vector of second-order exogenous latent variables explaining variance in $\boldsymbol{\eta}$, Γ is a matrix of partial regression coefficients for the regression of $\boldsymbol{\eta}$ on ξ, and ζ is a vector of errors in equations for $\boldsymbol{\eta}$. The implied variance–covariance matrix for the general SCFA model is

$$\Sigma = \begin{pmatrix} \Lambda_y(\Gamma\Phi\Gamma' + \Psi)\Lambda_y' + \Theta_\varepsilon & \Lambda_y\Gamma\Phi \\ \Phi\Gamma'\Lambda'_y & \Phi \end{pmatrix} \tag{6}$$

where Θ_ε is the covariance matrix for $\boldsymbol{\varepsilon}$ (which we assume is diagonal for now), Φ is the covariance matrix for ξ, and Ψ is the covariance matrix for ζ (also assumed diagonal for now). The decomposition of variance for **y** into common variance, specific variance, and random error variance can be computed, respectively, from $\hat{\Lambda}_y\hat{\Gamma}\hat{\Phi}\hat{\Gamma}'\hat{\Lambda}_y'$, $\hat{\Lambda}_y\Psi\hat{\Lambda}_y'$, and $\hat{\Theta}_\varepsilon$. Common variance in the measures refers to that accounted for from a particular ξ, and specific variance refers to that explained by a particular η exclusive of ξ. Later we introduce somewhat different notions of trait variance and measure specificity in more complex models of construct validity.

When one has a single second-order latent variable, at least three first-order latent variables are needed for identification. With only two first-order latent variables per second-order variable, identification can be achieved by fixing $\gamma_1 = \gamma_2$. This constraint hypothesizes that the contributions of the first-order latent variables, as indicators of the second-order latent variable, are equal. Such a constraint might be reasonable when first-order variables are equally important as indicators of a second-order factor. Under certain conditions, first-order latent variables can be modeled as functions of each other (for an example, see Yi, 1989). Bagozzi, 1994b, pp. 339–341) presents an example of a SCFA (see also Bagozzi, 1982).

1.3. Panel models and measure specificity

Fig. 1 illustrates a panel model that can be used to examine reliability and construct validity, while at the same time permitting partitioning of measure

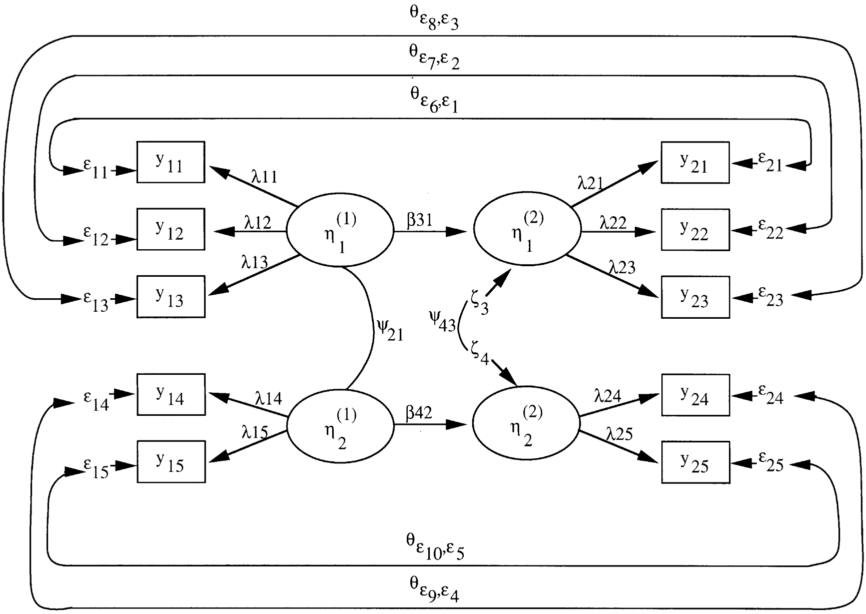

Fig. 1. Structural equation panel model for examining reliability and construct validity of two constructs at two points in time.

variance into trait, measure specific, and random error components. In the model, two constructs, η_1 and η_2, are shown measured by three and two indicators, respectively, at each of two points in time. The temporal stabilities of η_1 and η_2 are shown by β_{31} and β_{42}, respectively, which reflect corrections for random and specific measurement errors, as developed shortly. The parameters, β_{31} and β_{42}, estimate test-retest reliabilities of the measures of η_1 and η_2.

Discriminant validity between measures of η_1 and η_2 can be assessed by inspection of Ψ_{21} and Ψ_{43}. Discriminant validity refers to the degree to which measures of different latent variables are distinct. Purported measures of different variables that are too highly associated (in relation to the association among measures for each of the variables) bring into question the basis for claiming that the variables are unique. Discriminant validity is especially a problem with variables in consumer research where many similar or related subjective states are typically modeled. To avoid circular reasoning in such contexts, discriminant validity must be demonstrated. Notice in Fig. 1 that Ψ_{21} is the correlation between the two latent variables measured at Time 1, whereas ψ_{43} is the partial covariation between the same two latent variables measured at Time 2 (i.e., it is the covariation after partialing out variance due to dependence on the same measures at Time 1).

Application of appropriate estimation procedures to the simultaneous equations implied by Fig. 1 will yield estimates for Ψ_{21} and Ψ_{43} that correct for random and measure specific errors. As an example of the basis for partitioning these errors, consider the measurement equations for the first measure of η_1 at both points in time:

$$y_{11} = \lambda_{11}\eta_1^{(1)} + \varepsilon_{11}, \tag{7}$$

$$y_{21} = \lambda_{21}\eta_1^{(2)} + \varepsilon_{21}. \tag{8}$$

Rewriting these equations to express separate effects for random and specific variance on the measures gives

$$y_{11} = \lambda_{11}\eta_1^{(1)} + s_{\mathrm{p}1} + e_{11} \tag{9}$$

$$y_{21} = \lambda_{21}\eta_1^{(2)} + s_{\mathrm{p}1} + e_{21} \tag{10}$$

where $s_{\mathrm{p}1}$ is the measure specific error and e_{11} and e_{21} are the random error components of y_{11} and y_{21}, respectively. Specific variance in each measure can thus be estimated by modeling serially correlated errors: e.g., var $(s_{\mathrm{p}1}) =$ cov $(\varepsilon_{11}, \varepsilon_{21}) = \theta_{\varepsilon 6,\varepsilon 1}$. As measure specificity is a consistent component of each measure, over and above that reflected by the corresponding latent variable, it will contribute to the reliability of measures. However, the portion of variance due to each latent variable, which is reflected in the λ_{ij}'s, is of greater theoretical interest, as it provides information on trait variance.

Related to this, convergent validity addresses the issue of how well do multiple measures of a latent variable agree in their assessment of the variable. This is reflected in the magnitude of the λ_{ij}'s in Fig. 1. Given a satisfactory goodness-of-fit for the entire model, the greater the estimates for the λ_{ij}'s, the stronger the convergent validity.

As a final comment on the model in Fig. 1, we should note that it is possible to introduce cross-lagged paths into the figure to represent the effects of $\eta_1^{(1)}$ on $\eta_2^{(2)}$ (i.e., β_{41}) and $\eta_2^{(1)}$ on $\eta_1^{(2)}$ (i.e., β_{32}). These paths reflect the causal influence, if any, between the variables over time and may be of substantive interest, depending on the context under investigation. Bagozzi and Yi (1993, pp. 158–165) provide an illustration of the panel model shown in Fig. 1 in a marketing context.

1.4. Modeling method variance as additive effects

The CFA, SCFA, and panel models do not take into account method variance. To the extent that method biases exist, these models will provide poor fits to data and yield misleading inferences.

One way to model method variance is to modify the specification underlying the CFA model. In particular, a general additive 'trait-method' (ATM) model can be expressed as

$$y = [\Lambda_T \quad \Lambda_M] \begin{bmatrix} \eta_T \\ \eta_M \end{bmatrix} + \varepsilon \tag{11}$$

$$\Sigma = \Lambda_T \Psi_T \Lambda_T' + \Lambda_M \Psi_M \Lambda_M' + \Theta \tag{12}$$

where y is a vector of rs observed measures for r traits and s methods, $\eta = [\eta_T \quad \eta_M]'$ is an $(r+s) \times 1$ vector of trait and method factors, ε is a vector of rs residuals for $\boldsymbol{y}$, Σ is an $rs \times rs$ implied covariance matrix for $\boldsymbol{y}$, Ψ_T is an $r \times r$ correlation matrix for traits, Ψ_M is an $s \times s$ correlation matrix for methods, Θ is an $rs \times rs$ diagonal matrix of unique variances for $\boldsymbol{\varepsilon}$, $\Lambda_T = [\Lambda_1, \Lambda_2', \ldots, \Lambda_s]'$, Λ_j is an $r \times r$ diagonal matrix with trait factor loadings for the r traits measured by the j-th method, and

$$\Lambda_M = \begin{bmatrix} \lambda_1 & \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \lambda_2 & \cdot & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \cdot & \mathbf{0} & \lambda_s \end{bmatrix} \tag{13}$$

where λ_j is an $r \times 1$ vector of factor loadings for the jth method and traits are assumed uncorrelated with methods. This model is sometimes termed an MTMM (multitrait-multimethod) model.

The major assets of the ATM model are the following. Unlike classic procedures for investigating construct validity which rest on computation and inspection of correlations (e.g., Campbell and Fiske, 1959), the overall fit of any ATM model can be assessed with a χ^2-test and other diagnostics. The ATM model provides explicit information, as well, on convergent and discriminant validity. The former is achieved when factor loadings in Λ_T are statistically significant and high in magnitude. The latter is attained when confidence intervals around correlations among traits in Ψ_T do not include 1.00. Another benefit of the ATM model is the ability to partition measure variance into trait, method, and error components. This information can be used to verify the reliability and validity of measure and identify poor measures that need to be replaced or revised.

A limitation of the ATM model is that, in practice, it sometimes yields ill-defined solutions, or solutions cannot be found in the sense that the iteration procedure during estimation fails to converge. By ill-defined solutions we mean negative error variances or other illogical parameter estimates. When ill-defined solutions arise, a possible cause is model misspecification, and the ATM model may not be appropriate (cf., Kiers et al., 1996). Simulations show that ill-defined solutions occur frequently for the ATM model (e.g., Marsh and Bailey, 1991). However, these simulations were performed only for models where each method was used to measure a single indicator per trait. When each method is used to operationalize at least two measures per trait, the incidence of ill-defined solutions is likely to decline. Ill-defined solutions frequently result from over-parameterized models, so increasing the number of measures per trait from each method should overcome this problem. However, this remedy may prove impractical in practice, in that it requires at least double the number of measures of the standard ATM model.

1.4.1. Illustration

We applied the general ATM model to data collected by Phillips (1981). Phillips (1981) asked executives at 506 wholesale distribution companies to provide information on the nature of supplier influence on their operations. Five traits were measured: perceived influence that the supplier had over the size of orders, the mix of orders, salesforce hiring policies, salesforce training policies, and territories served. Information provided by the chief executive officer (CEO) in each company was treated as one method; while information provided by either another subsubordinate (in those companies where only a single additional informant was available) or the average of responses by two or more subordinates (in those companies where two or more additional informants were available) was treated as a second method. Fig. 2 shows the model.

The data fit the 5-trait, 2-method ATM model well: $\chi^2(14) = 13.76, \boldsymbol{p} = 0.47$. Table 1 presents the decomposition in variance due to trait, method, and error. It can be seen that high amounts of trait variance are explained by the CEO for supplier influence over the size and mix of orders, salesforce training policies, and territories served. The CEO was a poor informant with regard to supplier influence over salesforce hiring (see low trait and high error variance in Table 1). By contrast, subordinates were relatively informative with regard to supplier influence over salesforce hiring and territories served but noninformative with respect to influence over size and mix of orders and salesforce training. Overall, method error was very low with the exception of the CEO's assessment of supplier influence over salesforce hiring, which was quite large. Error variance was generally low for the CEO and moderate for subordinates.

The intercorrelations among traits and among methods are also shown in Table 1. Despite the subjectivity of measures and their apparent similarity, the correlations are all very low to moderate among traits, thereby demonstrating discriminant validity. The methods are correlated at a low level as well and thus show quite a bit of independence.

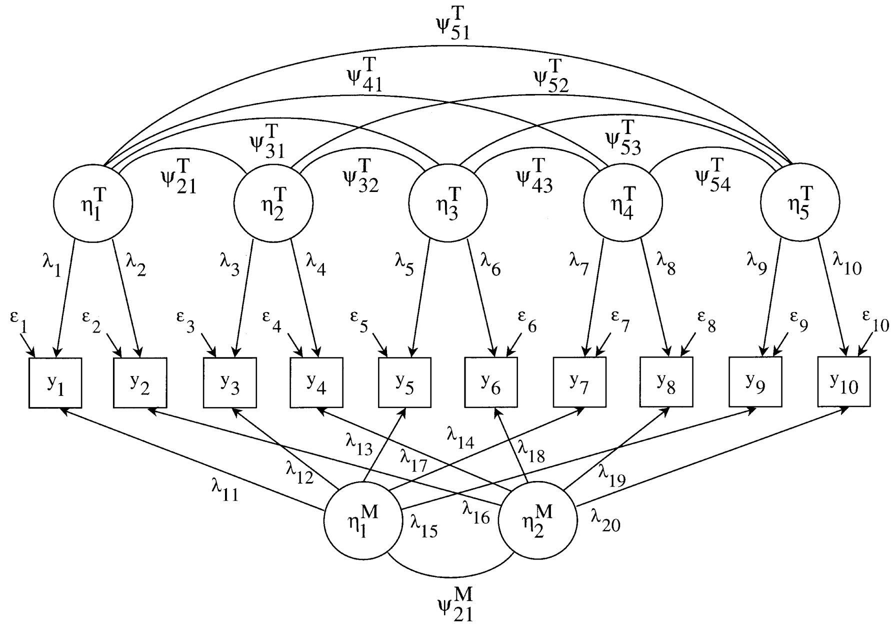

Fig. 2. The additive trait-metod-error confirmatory factor analysis model applied to five traits and two methods.

1.5. Modeling method variance with the correlated uniqueness model

An approach that is less prone to ill-defined solutions, yet accomplishes most of the objectives of the ATM model, is the correlated uniqueness (CU) model (Marsh, 1989). The CU model is the same as the CFA model presented in Eqs. (1) and (2), but with one modification. Instead of uncorrelated residuals as in the CFA model, the error terms for measures are permitted to be correlated within each method.

Two drawbacks with the CU model can be identified. First, the interpretation of covariances among errors will be problematic if, within the same method, significant positive and negative covariances result. It is difficult also to conceive of reasons why the same method could have opposite effects on measures of different traits when the traits are expected to covary in *either* a positive or negative direction. This incongruous outcome has occurred in some investigations (see Bagozzi and Yi, 1993, p. 150). Second, the CU model assumes that method factors are uncorrelated, a restriction often violated in practice. Bagozzi and Yi (1993) provide an illustration of the CU model.

Table 1
Decompostion of variance in measures

a. Additive trait-method model applied to study of supplier influence over the operations of wholesale distributors

Measure	Percentage variance due to		
	Trait	Method	Error
CEO			
Size of orders (y_1)	0.81	0.02	0.17
Mix of orders (y_2)	0.93	0.01	0.06
Salesforce hiring (y_3)	0.14	0.65	0.21
Salesforce training (y_4)	0.62	0.13	0.25
Territories served (y_5)	0.64	0.06	0.29
Subordinate(s)			
Size of orders (y_6)	0.27	0.09	0.64
Mix of orders (y_7)	0.18	0.09	0.73
Salesforce hiring (y_8)	0.55	0.28	0.18
Salesforce training (y_9)	0.22	0.24	0.54
Territories served (y_{10})	0.52	0.06	0.43

b. Trait and method correlation matrices

$$\psi_T = \begin{bmatrix} 1.00 & & & & \\ 0.32(0.07)^a & 1.00 & & & \\ 0.11(0.08) & 0.11(0.07) & 1.00 & & \\ 0.18(0.07) & 0.20(0.07) & 0.08(0.10) & 1.00 & \\ 0.09(0.06) & 0.16(0.06) & 0.39(0.10) & 0.15(0.09) & 1.00 \end{bmatrix}$$

$$\psi_M = \begin{bmatrix} 1.00 & \\ 0.26(0.13) & 1.00 \end{bmatrix}$$

[a]Standard errors in parentheses.

1.6. Modeling method variance with covariance components analysis

A new approach to the investigation of construct validity has recently been proposed by Wothke (1995) and is called covariance components analysis (CCA). The CCA model is an adaptation of a multivariate random model for factorial measurement designs. Analogous to ANOVA models, each trait measured by each method is written as a linear function of a general term, trait effects, method effects, and residual error. For a set of measures $\boldsymbol{X}$, the latent structure decomposition can be written as

$$\boldsymbol{X} = \Xi \boldsymbol{A}' + \boldsymbol{E} \tag{14}$$

where $\boldsymbol{X}$ is an $n \times mt$ matrix of n units of measurement (e.g., people, organizations) and m measures times t traits, Ξ is an $n \times (1 + t + m)$ matrix of parameters representing a general variate ξ_{ig} (which reflects the general score on all measures for a unit of analysis), a trait vector ξ_{iT_j}, and a method vector ξ_{iM_k}; and $\boldsymbol{E}$ is an $n \times mt$ matrix of error components (which is assumed to be diagonal). The matrix $\boldsymbol{A}$ is an $(1 + t + m) \times mt$ structural coefficient matrix of 1's and 0's. The implied covariance matrix is

$$\Sigma = \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{A}' + \Theta, \tag{15}$$

where Φ is the covariance matrix for Ξ and Θ is a covariance matrix for the errors.

Not all parameters in Eq. (15) will be estimated, and to achieve identification a number of simplifications need to be made. One meaningful set of constraints, analogous to ANOVA designs, is to redefine the effects of traits and methods as contrasts. For example, Wothke (1995, p. 128) proposes that the covariance matrix of the rows of $\boldsymbol{X}$ can be expressed as

$$\Sigma = \boldsymbol{K}\boldsymbol{\Phi}\boldsymbol{K}' + \Theta \tag{16}$$

where $\boldsymbol{K}$ is a $mt \times (t + m - 1)$ matrix of meaningful contrasts. The nature of the contrast matrix will be clear when we present the example below. One other constraint is needed for identification. The variance of the general variate should be fixed to unity, which means that the estimated variances of traits and methods must be evaluated in relation to this variance.

A final specification issue concerns the metrics of the measures. Wothke (1995, p. 130) proposes use of scale-factor matrices to achieve a scale-free extension of the CCA model. With this revision, a version of the CCA model useful for many construct validity designs becomes

$$\Sigma = \boldsymbol{D}\boldsymbol{K}\boldsymbol{\Phi}\boldsymbol{K}'\boldsymbol{D} + \Theta \tag{17}$$

where D is a $mt \times mt$ diagonal matrix of scale parameters.

A number of variations of the CCA model might be explored further, depending on the purposes of the researcher. For example, in addition to the diagonal covariance component matrix Φ which we investigate below, two models of interest are one where correlations among traits and among methods are also hypothesized and one where these plus correlations between traits and methods are also hypothesized. In addition to planned contrasts as illustrated below, one may perform arbitrary contrasts if traits do not lend themselves to natural comparisons. In the interest of brevity, we do not present the above variations in analyses below.

The CCA model represents an interesting approach to construct validity, but we know little about its limitations because it is so new. Interestingly when no method effects are modeled under the CCA model, the resulting ‘trait-only’ version is identical to the CFA model. This property is also true of the ATM model, when no method effects are modeled. The CCA and ATM models are of course different from each other in their full trait plus method forms. Wothke (Wothke, 1995, p. 138) notes that the CCA model and the direct product model (which we will consider later in the article) have the same number of degrees of freedom and typically yield similar results, when the correlations among measures across and within methods are positive.

1.6.1. Illustration

To illustrate the CCA model, we applied it to measures of attitudes toward church studied by Ostrom (1969). In this study, affective, behavioral, and cognitive components of attitude constituted the traits and were each measured by Likert, Guttman, and self-rating scales, which served as the methods. The sample size was 189. For this 3-trait, 3-method model, the contrast matrix is specified as

$$\boldsymbol{K} = \begin{bmatrix} 1/3 & \sqrt{2}/3 & 0 & \sqrt{2}/3 & 0 \\ 1/3 & -1/\sqrt{18} & 1/\sqrt{6} & \sqrt{2}/3 & 0 \\ 1/3 & -1/\sqrt{18} & -1/\sqrt{6} & \sqrt{2}/3 & 0 \\ 1/3 & \sqrt{2}/3 & 0 & -1/\sqrt{18} & 1/\sqrt{6} \\ 1/3 & -1/\sqrt{18} & 1/\sqrt{6} & -1/\sqrt{18} & 1/\sqrt{6} \\ 1/3 & -1/\sqrt{18} & -1/\sqrt{6} & -1/\sqrt{18} & 1/\sqrt{6} \\ 1/3 & \sqrt{2}/3 & 0 & -1/\sqrt{18} & -1/\sqrt{6} \\ 1/3 & -1/\sqrt{18} & 1/\sqrt{6} & -1/\sqrt{18} & -1/\sqrt{6} \\ 1/3 & -1/\sqrt{18} & -1/\sqrt{6} & -1/\sqrt{18} & -1/\sqrt{6} \end{bmatrix}$$

The first column in $\boldsymbol{K}$ shows the constant weight for the general variate. Column two presents the contrast between Trait 1 and the average of Traits 2 and 3.

Column three addresses the contrast between Traits 2 and 3. The fourth and fifth columns provide contrasts for method-specific variance (Wothke, 1995, p. 129).

Table 2 summarizes the findings for the application of the CCA model to Ostrom's (Ostrom, 1969) data. This model yielded a $\chi^2(23) = 43.53$, $\boldsymbol{p} < 0.01$, RMSEA $= 0.07$, and CFI $= 0.98$. Thus on balance the fit is not too bad. The scale factors show relatively little variance and the error variances indicate that the Likert items are generally the best, the Guttman the worst, with the self-rating in the middle. The covariance components present the relative variances of the trait and method profiles, with the general variate variance standardized at one. As the relative trait variance for the first trait is nonsignificant ($\hat{\phi}_{1,r} = 0.008$, s.e. $= 0.006$), we can conclude that affect is not differentiated from behavior and cognition. However, $\phi_{T_2T_3r} = 0.020$ (s.e. $= 0.007$) shows that behavior and cognition are distinct. As the method variance components are both significant and larger than trait variance components, one may conclude that the method effects are substantial and preclude strong conclusions as to establishment of construct validity.

1.7. Partitioning variance into trait, method, specific, and error variance

All the procedures described above, which model method variance, confound random error with measure specific variance. Kumar and Dillon (1990) proposed the following model to address this issue:

$$Y_{ijk} = \lambda_{ijk}^{t}\eta_{i}^{t} + \lambda_{ijk}^{m}\eta_{j}^{m} + \lambda_{ijk}^{s}\eta_{k}^{s} + \varepsilon_{ijk} \tag{18}$$

where Y_{ijk} is the measure of the ith trait by the jth method on the kth item, ξ_i^t is a latent variable for the ith trait, η_j^m is a latent variable for the jth method, η_k^s is a latent variable for measure specificity of the kth item, the λ's are corresponding factor loadings, and ε_{ijk} is random error in Y_{ijk}. This model is similar to the ATM model shown in Eqs. (11) and (12) except for the error term which is replaced with $(\lambda_{ijk}^{s}\eta_{k}^{s} + \varepsilon_{ijk})$.

The trait–method–specific-error (TMSE) model provides detailed information on construct validity but has two drawbacks. The first is practical. Compared to the ATM model, at least twice the number of measures must be obtained to partition variation under the TMSE model. This may not be feasible in many studies. Second, because many factors are fit to the data under the TMSE model, the possibility of lack of convergence and/or ill-defined solutions may be great. Indeed, in our reanalyses of Phillip's (1980) data that are given below, considerable trial and error were required to provide suitable starting values for the iterative procedure to converge. We suspect that at least three measures per trait from each method will be needed for a viable TMSE model. This, of course, triples the number of measures, over and above the ATM model,

Table 2
Results for covariance components analysis applied to Ostrom's (1969) attitude data

Variable/method	Sacle $\hat{D}$	Fixed contrasts K					Error variance $\hat{\Theta}$
		K_g	K_{T1}	K_{T2}	K_{M1}	K_{M2}	
Affect/Likert	2.55(0.18)[a]	0.33	0.47	0.00	0.47	0.00	0.22(0.03)
Behavior/Likert	2.66(0.17)	0.33	− 0.24	0.41	0.47	0.00	0.13(0.03)
Cognition/Likert	2.61(0.17)	0.33	− 0.24	− 0.41	0.47	0.00	0.16(0.03)
Affect/Guttman	1.83(0.20)	0.33	0.47	0.00	− 0.24	0.41	0.59(0.07)
Behavior/Guttman	2.17(0.19)	0.33	− 0.26	0.41	− 0.24	0.41	0.42(0.05)
Cognition/Guttman	2.35(0.19)	0.33	− 0.26	− 0.41	− 0.24	0.41	0.31(0.05)
Affect/self	2.67(0.17)	0.33	0.47	0.00	− 0.24	− 0.41	0.13(0.03)
Behavior/self	2.36(0.18)	0.33	− 0.26	0.41	− 0.24	− 0.41	0.31(0.04)
Cognition/self	2.40(0.18)	0.33	− 0.26	− 0.41	− 0.24	− 0.41	0.29(0.04)
	Covariance components $\hat{\Phi}$						
	1						
	0	0.008(0.006)					
	0	0	0.020(0.007)				
	0	0	0	0.035(0.008)			
	0	0	0	0	0.043(0.012)		

[a]Standard errors in parentheses.

and makes the approach less practical. To obtain proper solutions for parameters with a TMSE model (and other models prone to overparameterization), it may be necessary further to constrain a small number of parameters (e.g., selected method loadings and/or error variances) to zero.

1.7.1. Illustration

As an illustration, we reanalyzed data reported in Phillips (1980). Phillips (1980) studied the use of the computer by 205 distribution companies in market forecasting, accounting, and financial practices. These three traits were measured by key informant reports provided by the CEO and a subordinate in each firm. Each informant provided two assessments of each trait, and thus there were 12 measures in total (the correlation matrix of the measures can be found in Bagozzi et al., 1991, p. 454). The goodness-of-fit for the model was $\chi^2(32) = 21.10$, $p = 0.93$, an excellent fit. Table 3 presents key parameter estimates. It can be seen that, based on the significant and moderate to high trait factor loadings, convergent validity is achieved. Likewise as revealed in the bottom of the table, discriminant validity between traits and between methods (i.e., informants) is demonstrated. The loadings in Table 3 can be used to express the relative contribution of latent variables to variance in measures. Variances due to latent variables can be computed by squaring the appropriate standardized loadings. Table 4 shows the variance decomposition, where it can be seen that trait variance is generally high except for two measures (Y_{121} and Y_{122}) on the first trait (use of the computer in market forecasting) as provided by the second method (a subordinate to the CEO in the firm). Note that Y_{121} shows moderately high method variance and Y_{122} reveals both moderate method and error variance. Random error is generally low across items, and item specific variance is even lower yet. These findings suggest that construct validity of measures is very good, yet point to two measures in need of improvement.

1.8. Modeling trait-method interactions

It has been suggested that method factors may interact with trait factors in a multiplicative way (e.g., Campbell and O'Connell, 1967). That is, the higher the relationship between traits, the higher the method effects. Until recently, no unambiguous method existed for representing this.

Browne (1984, 1989) proposed the following direct product (DP) model for representing the interaction between traits and methods (see also Cudeck, 1988):

$$\Sigma = Z(P_M \otimes P_T + E)Z \tag{19}$$

where Z is a nonnegative definite diagonal matrix of scale constants, some of which are set equal to unity to achieve identification, P_M and P_T are nonnegative definite method and trait correlation matrices, respectively, whose elements are

Table 3
Parameter estimates for model of trait, method, measure specific, and error variance of data reported in Phillips (1980)[a]

	Traits			Informants		Measure specificity						
Measure	T_1	T_2	T_3	I_1	I_2	S_1	S_2	S_3	S_4	S_5	S_6	Error
Y_{111}	0.74(0.10)			0.27(0.11)		0.39(0.08)						0.22(0.08)
Y_{112}	0.75(0.10)			0.35(0.11)			0.08(0.41)					0.30(0.08)
Y_{121}	0.40(0.11)				0.73(0.08)	0.39(0.08)						0.17(0.08)
Y_{122}	0.44(0.10)				0.60(0.08)		0.08(0.41)					0.45(0.07)
Y_{211}		0.77(0.09)		0.32(0.07)				0.29(0.16)				0.21(0.04)
Y_{212}		0.77(0.09)		0.31(0.06)					0.48(0.09)			0.06(0.03)
Y_{221}		0.77(0.09)			0.34(07)			0.29(0.16)				0.20(0.04)
Y_{222}		0.66(0.08)			0.36(0.06)				0.48(0.09)			0.18(0.03)
Y_{311}			0.72(0.09)	0.59(0.08)						0.08(0.44)		0.15(0.04)
Y_{312}			0.73(0.08)	0.54(0.07)							0.21(0.18)	0.16(0.04)
Y_{321}			0.79(0.08)		0.44(0.09)					0.08(0.44)		0.18(0.03)
Y_{322}			0.83(0.08)		0.41(0.09)						0.21(0.18)	0.11(0.03)
		Trait correlations						Informant correlations				
		1.00						1.00				
		0.40(0.09)		1.00				0.34(0.16)		1.00		
		0.31(0.10)		0.31(0.09)		1.00						

[a]Standard errors are in parentheses. Y_{ijk} is the measure for the ith trait by the jth informant on the kth item.

Table 4
Partioning of variance for model of trait, method, measure specific, and error variance of data reported in Phillips (1980)[a]

Measure	Trait variance	Method variance	Item specific variance	Error variance
Y_{111}	0.56	0.08	0.15	0.22
Y_{112}	0.57	0.13	0.01	0.30
Y_{121}	0.16	0.53	0.15	0.17
Y_{122}	0.19	0.36	0.01	0.45
Y_{211}	0.60	0.10	0.09	0.21
Y_{212}	0.60	0.10	0.24	0.06
Y_{221}	0.59	0.11	0.09	0.20
Y_{222}	0.43	0.13	0.24	0.18
Y_{311}	0.52	0.35	0.01	0.15
Y_{312}	0.53	0.30	0.04	0.16
Y_{321}	0.63	0.19	0.01	0.18
Y_{322}	0.68	0.17	0.04	0.11

[a] Y_{ijk} is the measure for the ith trait by the jth informant on the kth item.

particular multiplicative components of common-score correlations (i.e., correlations corrected for attenuation), E is a diagonal matrix of nonnegative unique variances, and $\otimes$ indicates a right direct (Kronecker) product. It can be seen that Eq. (19) decomposes measures into latent variable plus error-score components. Under Eq. (19) the correlation matrix corrected for attenuation has a direct product structure,

$$P_C = P_M \otimes P_T \tag{20}$$

where P_C is the disattenuated correlation matrix with a typical element $\rho(T_iM_k, T_jM_l)$, P_M is the latent method correlation matrix with a typical element $\rho(M_k, M_l)$, and P_T is the latent trait correlation matrix with a typical element $\rho(T_i, T_j)$. From the definition of a right direct product, one can see that a typical element of Eq. (20) is

$$\rho(T_iM_k, T_jM_l) = \rho(T_i, T_j)\rho(M_k, M_l). \tag{21}$$

Notice that this equation assumes a multiplicative structure for latent variables, rather than for measurements.

Wothke and Browne (1990) have recently shown that the DP model can be reformulated as a linear model. Because their presentation is rather elliptical, and the DP model serves as a foundation for the three-facet model introduced at the end of the article, we will develop the structural equation representation in some detail. Specifically, Eq. (19) can be written as a second-order confirmatory factor analysis model as follows:

$$\Sigma = \Lambda \Gamma \Phi \Gamma' \Lambda' \tag{22}$$

where $\Lambda = Z$, Γ is the partitioned matrix,

$$\Gamma = (C_M \otimes I_t | I_{mt}) = (\Gamma_1 | \Gamma_2) \tag{23}$$

C_M is a square, lower triangular matrix chosen such that $P_M = C_M C'_M$, I_t and I_{mt} are identity matrices, and

$$\Phi = \left(\begin{array}{c|c} I_m \otimes P_T & O \\ O & E \end{array} \right) = \left(\begin{array}{c|c} \Phi_1 & O \\ O & \Phi_2 \end{array} \right). \tag{24}$$

The DP model can be easily restricted to suitable submodels. One useful version of the model, a composite-error model, is defined by the additional restriction

$$E = E_M \otimes E_T \tag{25}$$

with E_M and E_T diagonal. By using the fact that any symmetric, nonnegative definite matrix can be expressed as the product of a square matrix and its transpose, this restriction can be rewritten as follows:

$$E = (E_M^{1/2} \otimes I_t)(I_m \otimes E_T)(E_M^{1/2} \otimes I_t)' = \Gamma_2 \Phi_2 \Gamma'_2. \tag{26}$$

Several restrictions are needed for the identification of the DP model. First, one equality constraint per method is required for identification of scale-factor estimates (Wothke and Browne, 1990). This restriction will fix the scale of the component scores. For instance, one may select a trait and set its scale factors in Z equal to unity. Alternatively, one can constrain all diagonal elements of C_M to unity. The two types of restrictions may be suitably combined. Another restriction is required in order to fix the scale of the error components, since $(aE_M) \otimes (bE_T) = E_M \otimes E_T$ for any $a = 1/b$. This may be achieved by fixing one element in either E_M or E_T at unity.

P_T is directly estimated in the model, and standard errors of its elements will be available from the LISREL solution. In contrast, the estimate of P_M is obtained by rescaling $C_M C'_M$ into a correlation matrix, and standard errors are not available from the LISREL output. However, one can obtain the standard errors for the method correlations by employing an alternative parameterization in which P_M (rather than P_T) is directly estimated. Alternatively, the MUTMUM program can be used (Browne, 1991).

Campbell and Fiske's (Campbell and Fiske, 1959) original criteria for convergent and discriminant validity have the following interpretations in the DP model (Browne, 1984, pp. 9,10). Evidence for convergent validity is achieved when the correlations among methods in P_M are positive and large. The first criterion for discriminant validity is met when the correlations among traits in P_T are less than unity. The second criterion for discriminant validity is attained when the method correlations in P_M are greater than the trait correlations in P_T. The final discriminant validity criterion is satisfied whenever the DP model holds. These interpretations follow from the DP model specification (Bagozzi and Yi, 1992).

The DP model is an innovative solution to the representation of interactions between traits and methods. However, three shortcomings should be noted. First, in comparison to the models described up to this point, the DP model provides a more global and less detailed standard for convergent validity. The requirement that method correlations be substantial is a composite indicator for convergence of multiple measures of each trait. The criterion does not supply information about the degree of convergence or point out which measure(s) is satisfactory or not. A related shortcoming of the DP model is that it is not possible to arrive at an estimate of variation in a measure due to traits, as is possible under the ATM model. Trait and method variance are confounded in the DP model.

Another problem is that, on occasion, the DP model fits the same data that the ATM, CU, or CCA models fit. When this has occurred, researchers have typically been able to rule in favor of one model based on the occurrence of ill-defined solutions or other criteria applied to the alternatives (e.g., for the ATM and CU models in such cases, see Bagozzi and Yi, 1991, p. 438). Nevertheless, little is known about the conditions leading to multiple models fitting the same data set satisfactorily.

A final issue to note is that formal procedures have been developed for testing a number of hypotheses of interest under the DP model. Bagozzi and Yi (1992) derive and illustrate hypotheses concerning methods, traits, and the error matrix. Five hypotheses on methods included: method correlations following a Toeplitz pattern, all methods equally correlated, all methods completely unrelated, all methods equivalent, and subsets of methods equivalent. Four hypotheses on traits were all traits equally correlated, all traits completely unrelated, all traits equivalent, and subsets of traits equivalent. Finally, five hypotheses on $\boldsymbol{E}$ in Eq. (19) include $\boldsymbol{E}$ is unrestricted, $\boldsymbol{E} = \boldsymbol{E}_M \otimes \boldsymbol{E}_T$, $\boldsymbol{E} = \boldsymbol{I} \otimes \boldsymbol{E}_T$ $\boldsymbol{E} = \boldsymbol{E}_M \otimes \boldsymbol{I}$, and all diagonal elements in $\boldsymbol{E}$ are the same, where I is the identity matrix. It is beyond the scope of this paper to consider these hypotheses, and the reader is referred to Bagozzi and Yi (1992).

1.8.1. Illustration

We applied the DP model to data reported in Menezes and Elbert (1979). Menezes and Elbert (1979) investigated retail store image, where three traits

Table 5
Parameter estimates for the direct product model analysis of the data in Menezes and Elbert (1979)

Trait method	Communalities	Error	Trait corrrelations					Method correlations				
			T_1	SE	T_2	SE	T_3	M_1	SE	M_2	SE	M_3
T_1M_1	0.91	0.09	1.00					1.00				
T_2M_1	0.90	0.10	0.77	0.03	1.00			0.91	0.02	1.00		
T_3M_1	0.94	0.06	0.45	0.05	0.49	0.05	1.00	0.89	0.02	0.90	0.02	1.00
T_1M_2	0.91	0.09										
T_2M_2	0.90	0.10										
T_3M_2	0.94	0.06										
T_1M_3	0.88	0.12										
T_3M_3	0.86	0.14										
T_3M_3	0.92	0.08										

Note: T_i = trait, M_j = method, SE = standard error.

(store appearance, products, and prices) were each measured with three methods (Likert, semantic differential, and stapel scales). The sample size was 250. Overall, the model fit well: $\chi^2(25) = 26.49$, $p = 0.37$, $RMSEA = 0.015$, and $CFI = 1.00$.

Table 5 presents parameter estimates for the DP model analysis. Because the method correlations P_M are large and positive (range: 0.89–0.91), we conclude that convergent validity has been achieved. Likewise, because the trait correlations P_T are less than unity (range: 0.45–0.49), the first criterion for discriminant validity is met. Similarly, since the method correlations are greater than trait correlations, the second criterion for discriminant validity is satisfied. Finally, given that the goodness-of-fit index points to an acceptable fit and the matrices of intertrait correlations have the same pattern, which ever method is used, the third criterion for discriminant validity is attained.

2. A three-facet multiplicative model

Up to this point, we have limited inquiry to two-facet designs, with traits and methods constituting the two facets. However, researchers sometimes have to deal with three facets. Given the frequency with which panel designs are conducted in marketing research, it would be useful to extend the representation of construct validity models to include a third facet, time. When one conducts a cross-national study, a third facet would take place. As an interesting case to scrutinize, we will examine the one positing a three-way interaction between occasions, methods, and traits. However, the model can be generalized to any three-facet design.

The three-facet multiplicative model is a straightforward extension of the DP model (Browne, 1984; Cudeck, 1988). The following model thus expresses the covariance matrix for a completely crossed design for occasions, methods, and traits as

$$\Sigma = \boldsymbol{Z}(\boldsymbol{P}_O \otimes \boldsymbol{P}_M \otimes \boldsymbol{P}_T + \boldsymbol{E})\boldsymbol{Z} \tag{27}$$

where $\boldsymbol{Z}$ is a nonnegative definite diagonal matrix of scale constants, $\boldsymbol{P}_O$, $\boldsymbol{P}_M$, and $\boldsymbol{P}_T$ are nonnegative definite occasion, method, and trait correlation matrices, respectively, whose elements are particular multiplicative components of correlations corrected for attenuation, and $\boldsymbol{E}$ is a diagonal matrix of nonnegative error variances.

2.1. Illustration

We collected data to illustrate the three-facet multiplicative model. One hundred and twenty three students participated in a study of weight loss/weight

Table 6
Parameter estimates for three-facet multiplicative model of consumer attitudes toward weight loss

Correlation matrices

Occasions, $\boldsymbol{P}_O$				Methods, $\boldsymbol{P}_M$			Methods, $\boldsymbol{P}_T$		
Time 1	1.00			Pleasant–unpleasant	1.00		Success attitude	1.00	
Time 2	0.67	1.00		Happy–unhappy	0.10	1.00	Process attitude	0.98	1.00
Time 3	0.65	0.77	1.00						

Measure[a]	Scale constraints, $\boldsymbol{Z}$	Error variance, $\boldsymbol{E}$
Y_{111}	0.92	0.09
Y_{112}	0.78	0.49
Y_{121}	1.05	0.37
Y_{122}	0.83	0.89
Y_{211}	0.80	0.13
Y_{212}	0.73	0.23
Y_{221}	0.97	0.45
Y_{222}	0.87	0.30
Y_{311}	0.74	0.18
Y_{312}	0.79	0.03
Y_{321}	1.09	0.10
Y_{322}	0.95	0.15

[a] Y_{ijk}, for $i = 1,2,3$ occasions, $j = 1, 2$ methods, $k = 1, 2$ traits.

maintenance. On each of three occasions separated by two weeks, subjects filled-out a questionnaire and had their body weight monitored. Two attitudes (traits) toward losing/maintaining one's body weight were measured: attitudes toward trying to lose weight and succeeding and attitudes toward the process needed to lose or maintain one's body weight. The former refers to an outcome or goal, the latter to the means used to achieve that goal. Bagozzi and Warshaw (1990) showed that these attitudes are instrumental in efforts to lose or maintain one's weight. As methods, two 7-point semantic differential items were used to measure attitudes: a pleasant-unpleasant scale and a happy-unhappy scale. As these are quite similar, our illustration is not a good one for showing distinct method effects. It is possible that pleasantness and happiness are relatively distinct for this particular action and the respondents.

All empirical analyses conducted in this article up to this point were performed with the LISREL8 program (Jöreskog and Sörbom, 1993). It is not possible at present to adapt LISREL8 for estimation of parameters and developing goodness-of-fit tests for the three-facet multiplicative model. As a result, we used a BMDP routine and FORTRAN code suggested by Cudeck (1988) to produce GLS estimates. This routine is based on a procedure proposed by Lee and Jenrich (1984) for the analysis of covariance structure models. The version used in this article is available from the authors on request.

The findings showed that the model fits satisfactorily: $\chi^2(49) = 66.35$, $p = 0.05$. Table 6 summarizes the GLS parameter estimates. The correlation matrix for occasion composites indicates that attitudes were fairly stable over the month-long study. By contrast, the methods show little association ($\hat{\rho}_M = 0.10$). Apparently subjects approached each measure as a unique method of assessment. Next, notice that the two traits are highly correlated: $\hat{\rho}_t = 0.98$ (recall that the correlations in $\boldsymbol{P}_O$, $\boldsymbol{P}_M$, and $\boldsymbol{P}_T$ are corrected for attenuation due to unreliability in the raw measures). Subjects do not differentiate between attitudes toward trying and succeeding and attitudes toward the process of trying. Except for attitudes toward the process measured at time 1 by happy–unhappy, all measures demonstrate low to moderate levels of error. In sum, from the standpoint of prediction for the particular subjects investigated in this study, it appears to make little difference whether one uses attitudes toward success or attitudes toward the process when predicting a criterion.

3. Guidelines for investigating construct validation

How should one conduct a construct validation study, given the many options? The answer to this depends on whether one has a weak or strong theory concerning how traits and measurement error function. With no or a weak basis for modeling trait and error, an exploratory procedure is recommended. The most intuitive model, the additive trait-method model, should be run first. A satisfactory fit here provides a basis for assessing convergent and discriminant validity and partitioning variance in measures into trait, method, and error components. If estimation of the model fails to converge or improper solutions arise, the correlated uniqueness model may be appropriate, as it retains the assumption of additive trait or method effects. Alternatively covariance components analysis might be tried at this stage. If none of the above proves satisfactory, the reason may be that traits and methods interact and thus the direct product model should be employed.

With a strong theory for how traits and measurement error function, one should directly turn to the appropriate model and follow a confirmatory philosophy. Here there are two general cases to discuss. For cases where a single measure is available for each trait-method combination, the additive trait-method model is appropriate when one has reason to expect that traits, methods, and error are additive. If this model fails, the correlated uniqueness model may be appropriate, but only if uncorrelated methods are defensible. When one has reason to expect trait-method interactions, the direct product model should be investigated. With a third factor expected to interact with traits and methods, the three-facet multiplicative model should be tried. For cases where two or more measures are available for each trait-method combination, either the second-order confirmatory factor analysis model or the model proposed by Kumar and Dillon (1990) could be used. Of course the more parsimonious additive trait-method, correlated uniqueness, and covariance components models could be adapted here, if appropriate.

For cases where time is treated as a method, a number of possibilities exist. The panel model described earlier might be useful, especially when one desires to partial-out measure specificity. Alternatively if differential attenuation in measures is expected to interact with traits, the direct product model could be investigated. And if time is a third factor expected to interact with traits and methods, the three-facet multiplicative model should be scrutinized directly.

Finally, a very new, promising model should be mentioned. Constrained components analysis is especially appropriate when one has reason to believe that traits and methods are additive, but the additive trait-method and correlated uniqueness models are inappropriate or fail to work in practice (Kiers et al., 1996). For well-behaved data where the additive effects of traits and methods are identified, the latter models perform well and the constrained components models is somewhat less useful because factor loadings (and hence conclusions as to convergent validity) will be biased upward. But when problems of improper solutions or empirical under identification arise, the constrained components model will not be affected adversely and may be a viable alternative.

4. Conclusion

This article has explored a number of models for modeling measurement error in economics and behavioral research. From the analyses conducted herein, it is apparent that measurement error is pervasive and often large in empirical research. We reviewed procedures that permit the researcher to detect and correct for random error, measure specificity, and systematic method biases. Most of the procedures assumed linear effects for the different sources of error. However, we also considered a trait $\times$ method procedure and a trait $\times$ method $\times$ occasion approach to measurement error. In addition to describing and illustrating various approaches, we made comparisons where appropriate and pointed out assumptions and limitations of the procedures. Perhaps more so than in the field of economics and other areas of business research, such as finance and accounting, marketing and behavioral research are subject to measurement error and in need of remedies. It is hoped that researchers will use the procedures considered herein to develop measures and test hypotheses so as to explicitly correct for measurement error and improve decision making and inferences.

Acknowledgements

The authors thank Dr. Robert Cudeck, Department of Psychology, University of Minnesota, for providing the MBDP routine and FORTRAN code which served as the basis for the GLS procedure used herein for the TFM model. Special thanks are expressed to the reviewers and editors for the many recommendations they made.

References

Arbuckle, J.L., 1995. Amos Users' Guide. SmallWaters, Chicago.

Bagozzi, R.P., 1982. A field investigation of causal relations among cognitions, affect, intentions, and behavior. Journal of Marketing Research 19, 562–584.

Bagozzi, R.P., 1994a. Measurement in marketing research: Basic principles of questionnaire design. In: Bagozzi, R.P. (Ed.), Principles of marketing research. Blackwell, Oxford, pp. 1–49.

Bagozzi, R.P., 1994b. Structural equation models in marketing research: Basic principles. In: Bagozzi, R.P. (Ed.), Principles of marketing research. Blackwell, Oxford, pp. 317–385.

Bagozzi, R.P., 1996. The role of arousal in the creation and control of the halo effect in attitude models. Psychology and Marketing 13, 235–264.

Bagozzi, R.P., Edwards, J.R., 1998. A general approach for representing constructs in organizational research. Organizational Research Methods 1, 45–87.

Bagozzi, R.P., Warshaw, P.R., 1990. Trying to consume, Journal of Consumer Research 17, 127–140.

Bagozzi, R.P., Yi, Y., 1991. Multitrait-multimethod matrices in consumer research. Journal of Consumer Research 17, 426–439.

Bagozzi, R.P., Yi, Y., 1992. Testing hypotheses about methods, traits, and communalities in the direct-product model. Applied Psychological Measurement 16, 373–380.

Bagozzi, R.P., Yi, Y., 1993. Multitrait-multimethod matrices in consumer research: Critique and new developments. Journal of Consumer Psychology 2, 143–170.

Bagozzi, R.P., Yi, Y., Phillips, L.W., 1991. Assessing construct validity in organizational research. Administrative Science Quarterly 36, 421–458.

Bearden, W.O., Rose, R.L., 1990. Attention to social comparison information: An individual difference factor affecting consumer conformity. Journal of Consumer Research 16, 461–471.

Bentler, P.M., 1990. Comparative fit indexes in structural models. Psychological Bulletin 107, 238–246.

Bentler, P.M., Wu, J.C.E., 1993. EQS/Windows User's Guide. BMDP Statistical Software, Los Angeles.

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Bone, P.F., Sharma, S., Shimp, T.A., 1989. A bootstrap procedure for evaluating goodness-of-fit indices of structural equation and confirmatory factor models. Journal of Marketing Research 26, 105–111.

Browne, M.W., 1984. The decomposition of multitrait-multimethod matrices. British Journal of Mathematical and Statistical Psychology 37, 1–21.

Browne, M.W., 1989. Relationships between an additive model and a multiplicative model for multitrait-multimethod matrices. In: Coppi, R., Bolasco, S. (Eds.), Multiway Data Analysis. North-Holland, Amsterdam, pp. 507–520.

Browne, M.W., 1991. MUTMUM: PC User's Guide. Department of Psychology, Ohio State University, Columbus.

Browne, M.W., Cudeck, R., 1993. Alternative ways of assessing model fit. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. Sage, Newbury Park, pp. 136–162.

Campbell, D.T., Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 56, 81–105.

Campbell, D.T., O'Connell, E.J., 1967. Methods factors in multitrait-multimethod matrices: Multiplicative rather than additive? Multivariate Behavioral Research 2, 409-426.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Rand McNally, Chicago.

Cudeck, R., 1988. Multiplicative models and MTMM matrices. Journal of Educational Statistics 13, 131–147.

Dodds, W.B., Monroe, K.B., Grewal, D., 1991. Effects of price, brand, and store information on buyers' product evaluations. Journal of Marketing Research 28, 307–319.

Fiske, D.W., 1982. Convergent-discriminant validation in measurements and research strategies. In: Brinberg, D., Kidder, L.H. (Eds.), Forms of Validity in Research. Jossey-Bass, San Francisco, pp. 77–92.

Foxall, G.R., Haskins, C.G., 1986. Cognitive style and consumer innovativeness. European Journal of Marketing 20, 63–80.

Ganster, D.C., Hennessey, H.W., Luthans, F., 1983. Social desirability response effects: Three alternative models. Academy of Management Journal 26, 321–331.

Johnson, M.D., 1988. Comparability and hierarchical processing in multialternative choice. Journal of Consumer Research 15, 303–314.

Jöreskog, K.G., 1969. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika 34, 183–202.

Jöreskog, K.G., Dag Sörbom, 1993. LISREL8 User's Reference Guide. Scientific Software International, Chicago.

Kiers, H.A.L., Yoshio, T., Ten Berge, J.M.F., 1996. The analysis of multitrait-multimethod matrices via constrained components analysis. Psychometricka 61, 601–628.

Kumar, A., Dillon, W.R., 1990. On the use of confirmatory measurement models in the analysis of multiple-informant reports. Journal of Marketing Research 27, 102–111.
Lee, S.-Y., Jennrich, R.I., 1984. The analysis of structural equation models by means of derivative-free nonlinear least squares. Psychometrika 49, 521–528.
Marsh, H.W., 1989. Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. Applied Psychological Measurement 13, 335–361.
Marsh, H.W., Bailey, M., 1991. Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. Applied Psychological Measurement 15, 47–70.
Marsh, H.W., Balla, J.R., Hau, K.-T., 1996. An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In: Marcoulides, G.A., Schumacker, R.E. (Eds.), Advanced Structural Equation Modeling. Lawrence Erlbaum, Mahwah, NJ, pp. 315–353.
Menezes, D., Elbert, N.F., 1979. Alternative semantic scaling formats for measuring store image. Journal of Marketing Research 16, 80–87.
Nicholls, J.G., Licht, B.G., Pearl, R.A., 1982. Some dangers of using personality questionnaires to study personality. Psychological Bulletin 92, 572–580.
Ostrom, T.M., 1969. The relationship between the affective, behavioral, and cognitive components of attitude. Journal of Experimental Social Psychology 5, 12–30.
Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1994. Reassessment of expectations as a comparison standard in measuring service quality: Implications for future research. Journal of Marketing 58, 111–124.
Phillips, L.W., 1980. On studying collective behavior: Methodological issues in the use of key informants, Unpublished Ph.D. Thesis, Northwestern University.
Phillips, L.W., 1981. Assessing measurement error in key informant reports: A methodological note on organizational analysis in marketing. Journal of Marketing Research 18, 395–415.
Phillips, L.W., Bagozzi, R.P., 1986. On measuring organizational properties of distribution channels: Methodological issues in the use of key informants. Research in Marketing 8, 313–369.
Rosenthal, R., Rosnow, R.L. (Eds.), 1969. Artifacts in Behavioral Research. Academic Press, New York.
Seidler, J., 1974. On using informants: A technique for collecting quantitative data and controlling measurement error in organizational analysis. American Sociological Review 39, 816–831.
Steenkamp, J.-B.E.M., Van Trijp, H.C.M., 1991. The use of LISREL in validating marketing constructs. International Journal of Research in Marketing 8, 283–299.
Tucker, L.R., Lewis, C., 1973. The reliability coefficient for maximum likelihood factor analysis. Psychometrika 38, 1–10.
Winkler, J.D., Kanouse, D.E., Ware, J.E. Jr., 1982. Controlling for acquiescence response set in scale development. Journal of Applied Psychology 67, 555–561.
Wothke, W., 1995. Covariance components analysis of the multitrait-multimethod matrix. In: Shrout, P.E., Fiske, S.T. (Eds.), Personality Research, Methods, and Theory. Erlbaum, Hillsdale, NJ, pp. 124–144.
Wothke, W., Browne, M.W., 1990. The direct product model for the MTMM matrix parameterized as a second order factor analysis model. Psychometrika 55, 255–262.
Yi, Y., 1989. An investigation of the structure of expectancy-value attitude and its implications. International Journal of Research in Marketing 6, 71–83.
Yi, Y., 1990. A critical review of consumer satisfaction. In: Zeithaml, V.A. (Ed.), Review of Marketing 1990. American Marketing Association, Chicago, pp. 68–123.
Zaichkowsky, J.L., 1985. Measuring the involvement construct. Journal of Consumer Research 12, 341–352.