On the reliability, consistency, and method-specificity based on the CT-C(*M*–1) model

Cherng G. Ding · Ten-Der Jane

Abstract In the present study, we discuss reliability, consistency, and method specificity based on the CT-C (*M*−1) model, which provides clear definitions of trait and method factors and can facilitate parameter estimation. Properties of the reliability coefficient, the consistency coefficient, and the method-specificity coefficient of the summated score for a trait factor are addressed. The consistency coefficient and the method-specificity coefficient are both functions of the number of items, the average item consistency, and the average item method specificity. The usefulness of the findings is demonstrated in an alternative approach proposed for scale reduction. The approach, taking into account both traits and methods, helps identify the items leading to the maximum of convergent validity or method effects. The approach, illustrated with a simulated data set, is recommended for scale development based on multitrait–multimethod designs.

Keywords Confirmatory factor analysis (CFA) ·
Consistency coefficient · CT-C(*M*−1) model ·
Method-specificity coefficient · Multitrait-multimethod
(MTMM) · Reliability coefficient · Scale reduction

Multitrait-multimethod (MTMM) analysis, proposed by Campbell and Fiske (1959), has become an essential strategy for examining the construct validity of psychological measures (Eid & Diener, 2006). In MTMM designs, each of several traits (constructs) is measured by each of several methods. For example, in Mount's (1984) study, the traits of administrative ability, ability to give feedback to subordinates, and consideration when dealing with others were measured by using the three methods: self-rating, supervisor rating, and subordinate rating. Convergent validity is supported when different methods yield consistent results in measuring the same trait. Discriminant validity is supported when the traits can be well distinguished from each other.

Confirmatory factor analysis (CFA) is the analytical approach that is most consistent with Campbell and Fiske's (1959) original formulation of the MTMM matrix (e.g., Lance, Noble, & Scullen, 2002; Schmitt & Stults, 1986; Widaman, 1985). The effects of all traits and all methods are examined simultaneously in MTMM–CFA models. Therefore, convergent and discriminant validities are assessed in the presence of method effects. Controlling for method effects can reduce potential bias in parameter estimates. Widaman proposed hierarchically nested CFA models for MTMM data to facilitate examining trait and method effects. Although trait effects reflect how well a measure can represent its underlying trait, method effects reflect how much it is affected by the measurement method. The amount of method effects differs greatly across disciplines (Cote & Buckley, 1987; Malhotra, Kim, & Patil, 2006; Mishra, 2000).

A well-known MTMM–CFA model is the correlated trait-correlated method (CT–CM) model, in which all traits to be studied are regarded as trait factors and all methods that are applied are regarded as method factors. The trait factors can covary among themselves, and the method factors can covary among themselves as well. However, trait and method factors are assumed to be uncorrelated. In addition, it is assumed that method factors do not interact

C. G. Ding (✉) · T.-D. Jane
Institute of Business and Management,
National Chiao Tung University,
118 Chung-Hsiao West Road, Section 1,
Taipei, Taiwan
e-mail: cding@mail.nctu.edu.tw

with trait factors. Although the CT–CM model is faithful to Campbell and Fiske's (1959) original theoretical formulation of the MTMM matrix, this model suffers from two disadvantages. First, estimation problems often occur, especially when method factors are correlated (e.g., Kenny & Kashy, 1992; Marsh, 1989; Marsh & Grayson, 1995), although using larger MTMM matrices (i.e., more traits and methods) together with larger samples may yield admissible solutions (Conway, Lievens, Scullen, & Lance, 2004; Marsh & Bailey, 1991). Second, neither the trait factors nor the method factors are clearly defined, making their interpretation difficult (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003; Pohl, Steyer, & Kraus, 2008). To improve the model, Eid (2000) and Eid et al. (2003) recommended the CT-C($M-1$) model. It is named CT-C ($M-1$) because the number of method factors is one less than the number of methods used. At first, one method, called the standard (or reference) method, has to be chosen as the comparison standard. Then, for each trait, the true score of an item measured by the standard method is taken as a predictor. The method variable for an item measured by a nonstandard method is the residual resulting from the regression of the item true score on the predictor. It is assumed that the method variables associated with the same nonstandard method are homogeneous, having a common method factor. The common method factor thus captures reliable residual effects that are specific to the items measured by that nonstandard method and are not shared with the items measured by the standard method. Three important consequences follow. The first one is the absence of the method factor for the standard method, and therefore the trait is confounded with the standard method. The second one is the uncorrelatedness of trait and method factors, allowing the estimation of variance components due to trait, method, and error effects for the items measured by nonstandard methods. The third one is that the method factors can be correlated. This correlation is a partial correlation. The CT-C($M-1$) model avoids the estimation problems inherent in the CT–CM model. However, the parameter estimates depend on the standard method chosen. The meaning of the trait factor varies with different standard methods. The choice of the standard method should be guided by the research question and substantive theory (Geiser, Eid, & Nussbeck, 2008).

Decomposition of the reliability coefficient into the consistency coefficient and the method-specificity coefficient based on MTMM models can be seen in the literature (e.g., Deinzer et al., 1995; Eid, 2000; Eid et al., 2003; Schmitt & Steyer, 1993). When trait factors, method factors, and errors are uncorrelated, the variance of an item can be decomposed into trait variance, method variance, and error variance. The reliability coefficient of the item is that part of the item variance explained by the trait and method factors. It is the part of variance that is not due to measurement error. The consistency coefficient represents the proportion of the item variance due to the trait factor. The method-specificity coefficient represents the proportion of the item variance due to the measurement method. The reliability coefficient is the sum of the consistency coefficient and the method-specificity coefficient. The consistency coefficient is a quantitative indicator of convergent validity (Eid 2000; Eid et al., 2003). Deinzer et al. (1995) mentioned that, the consistency coefficient rather than the reliability coefficient should be used for examining trait effects.

It is usually of interest in behavior studies to assess reliability of the sum of items in a scale. Analyses of reliability, consistency, and method specificity for the sum of items across different methods based on an MTMM design become more complex. As can be seen later, the reliability of the summated score for a trait factor is a joint function of the average item reliability and scale length (the number of items used to measure the trait factor), regardless of whether there exist method factors in the model. However, the consistency coefficient and the method-specificity coefficient—the two components of the reliability coefficient—of the summated score possess somewhat different properties, which are important when assessing the quality of a scale under MTMM designs. Relevant discussions seem not addressed in the literature. The purpose of the present study was to fill this gap. The findings will be useful for behavior researchers conducting MTMM studies.

Reliability as a function of the number of items

In classical test theory (CTT), $x_{ij} = t_{ij} + \varepsilon_{ij}$, $i = 1, ..., I$; $j = 1, ..., J_i$, where x_{ij} denotes the observed score on the jth item for the ith trait factor, t_{ij} is its true score, ε_{ij} is the corresponding error, I is the number of trait factors in the model, and J_i is the number of items for measuring the ith trait factor. For congeneric items, true scores on different items for a trait factor are linearly related. Centered by its mean μ_{ij}, item x_{ij} can be formulated by CFA as (McDonald, 1999, p. 78)

$$x_{ij} - \mu_{ij} = \lambda_{F_{ij}} F_i + \varepsilon_{ij}, \ i = 1, ..., I; j = 1, ..., J_i, \quad (1)$$

where F_i denotes the latent factor for trait i, and $\lambda_{F_{ij}}$ is the trait loading of x_{ij} on F_i. $\lambda_{F_{ij}}$ is the covariance of x_{ij} and F_i. It is assumed that $\sigma_{F_i \varepsilon_{ij}} = 0$ and $\sigma_{\varepsilon_{ij} \varepsilon_{i'j'}} = 0 (i \neq i'$ or $j \neq j')$.

For identification reasons, we usually set $\sigma_{F_i}^2 = 1$ or $\lambda_{F_{i1}} = 1$. The reliability of the summated score $X_i = \sum_{j=1}^{J_i} x_{ij}$ for trait F_i, denoted by rel_{F_i} and defined as $rel_{F_i} = \sigma_{T_i}^2 / \sigma_{X_i}^2$, where $T_i = \sum_{j=1}^{J_i} t_{ij} = \sum_{j=1}^{J_i} \lambda_{ij} F_i$, is then given by (Lord & Novick, 1968, Chap. 9)

$$rel_{F_i} = \frac{\left(\sum_{j=1}^{J_i} \lambda_{F_{ij}}\right)^2}{\left(\sum_{j=1}^{J_i} \lambda_{F_{ij}}\right)^2 + \sum_{j=1}^{J_i} \sigma_{\varepsilon_{ij}}^2}. \tag{2}$$

rel_{F_i} is the proportion of the variance of X_i explained by F_i, and can be reexpressed as

$$rel_{F_i} = \frac{J_i \overline{\theta}_{F_i}}{1 + (J_i - 1)\overline{\theta}_{F_i}}, \tag{3}$$

where $\overline{\theta}_{F_i} = \frac{(\overline{\lambda}_{F_i})^2}{(\overline{\lambda}_{F_i})^2 + \overline{\sigma}_{\varepsilon_i}^2}$, $\overline{\lambda}_{F_i} = \sum_{j=1}^{J_i} \lambda_{F_{ij}} / J_i$, $\overline{\sigma}_{\varepsilon_i}^2 = \sum_{j=1}^{J_i} \sigma_{\varepsilon_{ij}}^2 / J_i$, and $0 \leq \overline{\theta}_{F_i} \leq 1$. The reliability of the individual item x_{ij} for trait F_i is $\theta_{x_{ij}} = \lambda_{F_{ij}}^2 / (\lambda_{F_{ij}}^2 + \sigma_{\varepsilon_{ij}}^2)$, $j = 1, \ldots, J_i$, which is the proportion of its variance explained by F_i. Rather than the average variance extracted, given by $\overline{\lambda_{F_i}^2} / \left(\overline{\lambda_{F_i}^2} + \overline{\sigma_{\varepsilon_i}^2}\right)$ (where $\overline{\lambda_{F_i}^2} = \sum_{j=1}^{J_i} \lambda_{F_{ij}}^2 / J_i$) (Fornell & Larcker, 1981), $\overline{\theta}_{F_i}$ is defined as the average item reliability for F_i (McDonald, 1999, p. 125), and is actually the reliability adjusted for scale length. rel_{F_i} is a joint function of $\overline{\theta}_{F_i}$ and J_i (McDonald, 1999, p. 124). For a fixed value of $\overline{\theta}_{F_i}$, rel_{F_i} increases as a function of J_i. The properties are similar in that coefficient alpha is a joint function of the average interrelatedness among items and the number of items (e.g., Cortina, 1993). An estimate of the item reliability $\theta_{x_{ij}}$, resulting from CFA, can serve as an index for assessing the internal quality of item j for trait i. A trait is "abstract" if it is reflected in a series of mental or physical activities (Bergkvist & Rossiter, 2007, p. 183). Scales for less abstract traits are usually composed of fewer items, with higher item reliabilities. On the other hand, for more abstract traits, more items with lower item reliabilities are needed. Although both can achieve a required level of reliability (e.g., .7), the former has fewer items but a greater $\overline{\theta}$ than the latter. When $\overline{\theta}$ is small, more items are needed to reach a desired reliability. Adding an item with lower item reliability will reduce $\overline{\theta}$, but may increase rel. Dropping an item with lower item reliability will increase $\overline{\theta}$, but may decrease rel.

Reliability, consistency, and method-specificity based on the CT-C(M−1) model

For MTMM data, items are affected not only by their underlying trait factors but also the measurement methods. Let x_{ijk} denote the observed score of the jth item for trait F_i with method M_k. The CFA model in Eq. 1 can be extended to formulate a CT–CM or a CT-C(M−1) model, in which every trait is measured using every method with either a single item or multiple items. In the CT–CM model, every item loads on a trait factor and on a method factor. In the CT-C(M−1) model, however, all items belonging to the standard method load exclusively on a trait factor. There is no method factor for these items (Eid et al., 2003; Geiser et al., 2008). If the first method ($k = 1$) is chosen as the standard method, the basic equations of the CT-C(M−1) model are as follows:

$$\begin{aligned} x_{ij1} &= t_{ij1} + \delta_{ij1} = \mu_{ij1} + \lambda_{F_{ij1}} F_{i\_1} + \delta_{ij1} \\ &\quad \text{(for the standard method } (k = 1)), \\ x_{ijk} &= t_{ijk} + \delta_{ijk} = \mu_{ijk} + \lambda_{F_{ijk}} F_{i\_1} + \lambda_{M_{ijk}} M_k + \delta_{ijk}, \\ &\quad i = 1, \ldots, I, \, j = 1, \ldots, J_{ik}, k = 2, \ldots K, \end{aligned} \tag{4}$$

where t_{ijk} denotes the true score of x_{ijk}; μ_{ijk} denotes the mean of x_{ijk}; F_{i_1} denotes the ith trait factor measured by the first method, chosen as the standard method; M_k ($k > 1$) denotes the kth method factor; $\lambda_{F_{ijk}}$ denotes the trait loading of x_{ijk} on F_{i_1}; $\lambda_{M_{ijk}}$ denotes the method loading of x_{ijk} on M_k ($k > 1$); δ_{ijk} denotes the corresponding error; and J_{ik} denotes the number of items for measuring F_{i_1} with M_k. There are different ways (e.g., by fixing the variances of F_{i_1} and M_k to 1 [Geiser et al., 2008] or by fixing one loading per factor to 1 [and freely estimating the factor variances] [Eid et al., 2003]) to identify the model. In the present study, the factor variances are fixed to 1 because loadings represent trait and method effects to be assessed, and one would not want to give fixed values for them. It is assumed that $E(F_{i\_1}) = E(M_k) = 0$, $\sigma_{F_{i\_1}\delta_{ijk}} = 0$, $\sigma_{M_k\delta_{ijk}} = 0$, and $\sigma_{\delta_{ijk}\delta_{i'j'k'}} = 0$ ($i \neq i'$ or $j \neq j'$ or $k \neq k'$). In addition, traits and methods do not interact. Trait factors are allowed to covary, and method factors are allowed to covary as well. Without loss of generality, a CT-C(M−1) model with the first method as the standard method for an MTMM design with three traits (F_2, F_3, F_4) and three methods is illustrated in Fig. 1. x_{211}, x_{212}, and x_{213} are the items for measuring trait F_2 by using methods 1, 2, and 3, respectively. x_{311}, x_{321}, x_{312}, x_{322}, x_{332}, x_{313}, and x_{323} are the items for measuring trait F_3 by using the first method for the first two items, the second method for the middle three items, and the third method for the last two items. x_{411}, x_{412}, x_{413}, and x_{423} are the items for measuring trait F_4 by using the first method for the first item, the second method for the second item, and the third method for the last two items. M_1 is missing in the figure because it

Fig. 1 The CT-C($M-1$) model with the first method as the standard method for an MTMM design with three traits and three methods. F_{i_1} (i = 2, 3, 4) denotes the ith trait factor specific to the standard method (the first method). M_k ($k = 2, 3$) denotes the kth method factor. There is no method factor for items x_{211}, x_{311}, x_{321}, and x_{411}, measured by the first method
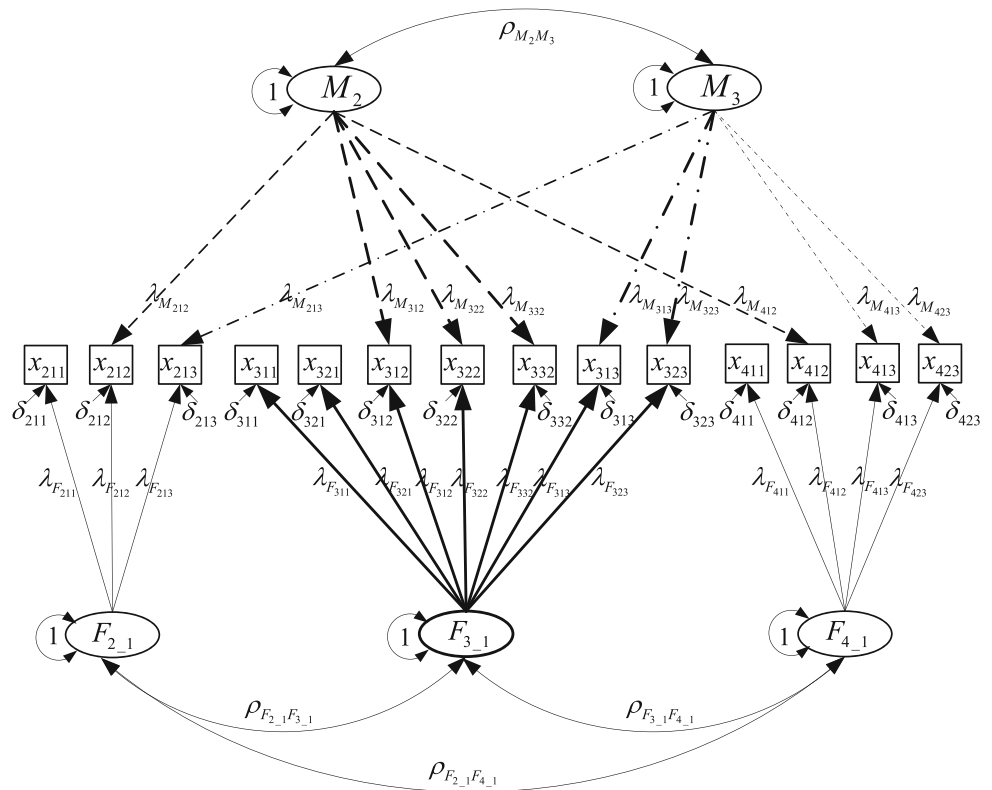
is used as the standard method. F_2, F_3, and F_4 are therefore the trait factors measured by the standard method (the first method). In the CT-C($M-1$) model, the three trait factors are better denoted by F_{2_1}, F_{3_1}, and F_{4_1} to make clear that they are specific to the first method.

Because the trait is confounded with the standard method, the corresponding items reflect both the trait and the standard method factors, and the trait and method effects are inseparable. Therefore, the consistency coefficient and the method-specificity coefficient cannot be obtained for the items belonging to the standard method. It follows that, when assessing consistency and method specificity for aggregated items, the items measured by the standard method need to be excluded. Under the assumptions for the CT-C($M-1$) model given in Eq. 4, the population reliability coefficient, consistency coefficient, and method-specificity coefficient of the summated score $\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}} x_{ijk}$ with respect to F_{i_1}— denoted respectively by $Rel_{F_{i\_1}}$, $CO_{F_{i\_1}}$, and $MS_{F_{i\_1}}$—are given by (see the Appendix for derivations)

$$Rel_{F_{i\_1}} = \frac{J_i \overline{\theta}^*_{F_{i\_1}}}{1 + (J_i - 1)\overline{\theta}^*_{F_{i\_1}}}, \tag{5}$$

$$CO_{F_{i\_1}} = \frac{J_i \overline{\omega}_{F_{i\_1}}}{1 + (J_i - 1)\overline{\theta}^*_{F_{i\_1}}}, \tag{6}$$

and

$$MS_{F_{i\_1}} = \frac{J_i \overline{\psi}_{F_{i\_1}}}{1 + (J_i - 1)\overline{\theta}^*_{F_{i\_1}}}, \tag{7}$$

w h e r e $J_i = \sum_{k=2}^{K} J_{ik}$, $\overline{\theta}^*_{F_i} = \overline{\omega}_{F_{i\_1}} + \overline{\psi}_{F_{i\_1}}$, $\overline{\omega}_{F_{i\_1}} = \frac{(\overline{\lambda}_{F_{i\_1}})^2}{(\overline{\lambda}_{F_{i\_1}})^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma}^2_{\delta_i}}$, $\overline{\psi}_{F_{i\_1}} = \frac{\Upsilon_{F_{i\_1}}}{(\overline{\lambda}_{F_{i\_1}})^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma}^2_{\delta_i}}$, $\Upsilon_{F_{i\_1}} = \sum_{k=2}^{K} (w_k \overline{\lambda}_{M_{ik}})^2 + \sum_{k \neq k'} (w_k \overline{\lambda}_{M_{ik}})(w_{k'} \overline{\lambda}_{M_{ik'}}) \rho_{M_k M_{k'}}$, $\overline{\lambda}_{F_{i\_1}} = \sum_{k=2}^{K} \sum_{j=1}^{J_{ik}} \lambda_{F_{ijk}}/J_i$, $\overline{\lambda}_{M_{ik}} = \sum_{j=1}^{J_{ik}} \lambda_{M_{ijk}}/J_{ik}$, $\overline{\sigma}^2_{\delta_i} = \sum_{k=1}^{K} \sum_{j=1}^{J_{ik}} \sigma^2_{\delta_{ijk}}/J_i$,

and $w_k = J_{ik}/J_i (k > 1)$. $\overline{\theta}^*_{F_{i\_1}}$ is an extended version of the average item reliability $\overline{\theta}_{F_i}$. It is a reliability index excluding the effects of scale length. $CO_{F_{i\_1}}$ indicates that part of the variance of the summated score (sum of the items measured with nonstandard methods) that is due to interindividual differences on F_{i_1}. $MS_{F_{i\_1}}$, on the other hand, represents that part of the variance that is due to method-specific differences (not shared with the standard method). $Rel_{F_{i\_1}} = CO_{F_{i\_1}} + MS_{F_{i\_1}}$. A large $CO_{F_{i\_1}}$ indicates high convergent validity. A large $MS_{F_{i\_1}}$ reflects high method effects and low convergent validity. Since Eq. 5 is the same as Eq. 3, the relevant properties discussed in the previous section for rel_{F_i} apply for $Rel_{F_{i\_1}}$.

$\overline{\omega}_{F_{i\_1}}$ and $\overline{\psi}_{F_{i\_1}}$, both independent of scale length, can be regarded as the average item consistency and the average item method-specificity, respectively. If $\lambda_{F_{ijk}} > \lambda_{M_{ijk}}$, then the jth item for F_{i_1} is affected more by the underlying trait factor than by the kth (nonstandard) method. If $\overline{\omega}_{F_{i\_1}} > \overline{\psi}_{F_{i\_1}}$, then the items used are affected more, on average, by the trait factor than by the nonstandard methods. In contrast, if $\lambda_{M_{ijk}} > \lambda_{F_{ijk}}$, then the influence of kth method on the item is greater than the influence of its underlying trait. $\overline{\psi}_{F_{i\_1}} > \overline{\omega}_{F_{i\_1}}$ indicates that the average method effect is greater.

By Eq. 6, CO is no longer a fixed function of $\overline{\omega}$ and the number of items J. Since $0 \leq \overline{\omega} \leq 1$ and $0 \leq \overline{\psi} \leq 1 - \overline{\omega}$, CO is bounded below by $\overline{\omega}$ when $\overline{\psi} = 1 - \overline{\omega}$, and is bounded above by $J\overline{\omega}/(1 + (J-1)\overline{\omega})$ (≤ 1) when $\overline{\psi} = 0$, denoted respectively by CO_L and CO_U. When $\overline{\omega} \leq .5$, if $\overline{\omega} \geq \overline{\psi}$, then $CO_M \leq CO \leq CO_U$, where $CO_M = J\overline{\omega}/(1 + 2(J-1)\overline{\omega})$, occurring at $\overline{\psi} = \overline{\omega}$; otherwise $CO_L \leq CO < CO_M$, as marked in Fig. 2 for $\overline{\omega} = .3$ and $J = 3$. For any fixed value of J, CO_L, CO_U, and CO_M are all increasing functions of $\overline{\omega}$. They are pictorially presented for $J = 3$ through 8 in Fig. 2, using a solid line, solid curves, and dashed curves for CO_L, CO_U and CO_M, respectively. Given a value of $\overline{\omega}$, CO_U is an increasing functions of J, and, if $\overline{\omega} < .5$, CO_M is also an increasing function of J. Given values of $\overline{\omega}$ and J, the less influence the methods exert, the closer CO is to CO_U. If $\overline{\omega} \leq \overline{\psi}$, $\overline{\omega}$ cannot exceed .5, and $CO \leq CO_M < .5$, regardless of the value of J. Under this case, using more items while keeping the same level of $\overline{\omega}$ cannot reach any level of CO greater than .5. Given a fixed level of CO, the number of items used could be reduced by increasing $\overline{\omega}$ and further

by decreasing $\overline{\psi}$. CO may be increased by eliminating the items with smaller trait loadings and larger method loadings, and/or by adding those with larger trait loadings and smaller method loadings.

By Eq. 7, MS possesses similar properties. MS is bounded below by $MS_L = \overline{\psi}$ when $\overline{\omega} = 1 - \overline{\psi}$, and is bounded above by $MS_U = J\overline{\psi}/(1 + (J-1)\overline{\psi})$ (≤ 1) when $\overline{\omega} = 0$. When $\overline{\psi} \leq .5$, if $\overline{\psi} \geq \overline{\omega}$, then $MS_M \leq MS \leq MS_U$, where $MS_M = J\overline{\psi}/(1 + 2(J-1)\overline{\psi})$, occurring at $\overline{\omega} = \overline{\psi}$; otherwise, $MS_L \leq MS < MS_M$. For any fixed value of J, MS_L, MS_U, and MS_M are all increasing functions of $\overline{\psi}$. They are similarly pictorially presented for $J = 3$ through 8 in Fig. 3. As shown in Fig. 3, given a value of $\overline{\psi}$, MS_U and MS_M are also increasing functions of J. Moreover, the less influence of the trait factor, the closer MS is to MS_U. However, MS cannot reach any level greater than .5 if $\overline{\psi} \leq \overline{\omega}$, regardless of the number of items used. Given a fixed level of MS, the number of items used could be reduced by increasing $\overline{\psi}$ and further by decreasing $\overline{\omega}$. The items with smaller method loadings and larger trait loadings could be considered to be eliminated to increase MS.

High convergent validity is not always the goal of research. Beyond the traditional search for maximum convergent validity, a thorough analysis of method influence might tell a different interesting story. Hence, a multimethod study should always have two facets: the examination of convergent validity and the analysis of method-specific influences (Eid & Diener, 2006, p. 5). When many items are selected for a trait to attain a required level of reliability, desirable convergent validity or the method effects may not be achieved with the same
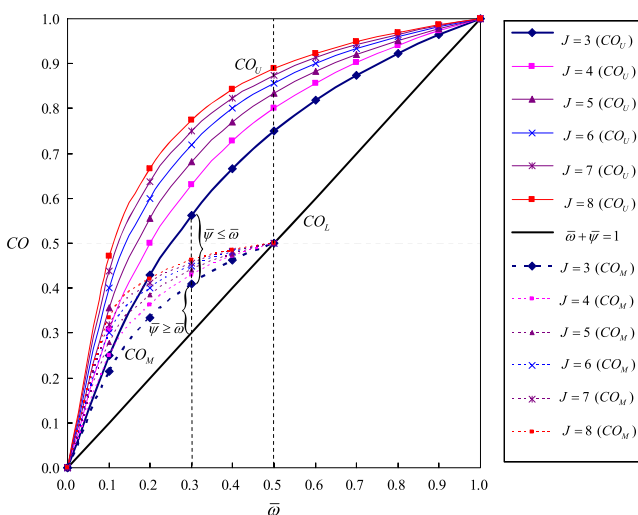
Fig. 2 CO (the consistency coefficient) as a function of $\overline{\psi}$ for a given value of J and a given value of $\overline{\omega}$. $CO_L = \overline{\omega}$, $CO_M = J\overline{\omega}/(1 + 2(J-1)\overline{\omega})$, $CO_U = J\overline{\omega}/(1 + (J-1)\overline{\omega})$
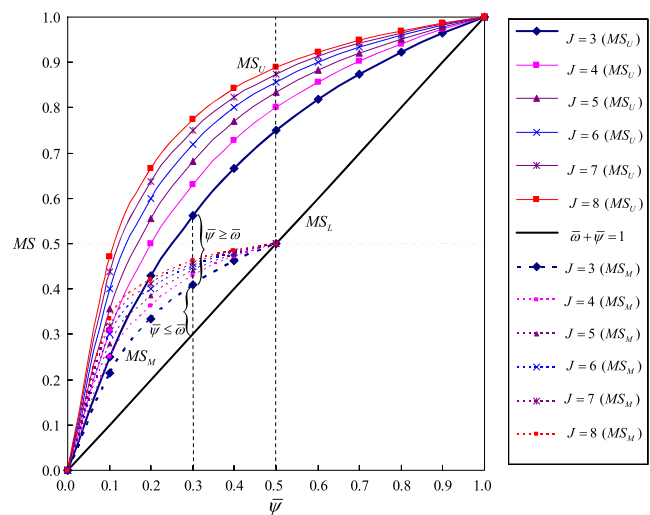
Fig. 3 MS (the method-specificity coefficient) as a function of $\overline{\omega}$ for a given value of J and a given value of $\overline{\psi}$. $MS_L = \overline{\psi}$, $MS_M = J\overline{\psi}/(1 + 2(J-1)\overline{\psi})$, $MS_U = J\overline{\psi}/(1 + (J-1)\overline{\psi})$

items. Further item selection is needed. As mentioned in Eid et al. ([2003](#), p. 47), in personnel selection, for example, items with a low degree of method specificity and a high degree of consistency are desirable when different raters rate applicants for a position. In marital therapy, on the other hand, those items with high method specificity and low consistency are of interest if the two different raters are spouses rating each other. In this context, psychologists might be particularly interested in the divergent views of the spouses.

Empirically, since $\widehat{Rel}$ (the sample Rel) $= \widehat{CO}$ (the sample CO) $+ \widehat{MS}$ (the sample MS), large values of $\widehat{Rel}$ may result from small values of $\widehat{CO}$ (reflecting poor convergent quality) but large values of $\widehat{MS}$, or from large values of $\widehat{CO}$ but small values of $\widehat{MS}$ (reflecting weak method influence). Acceptable reliability does not necessarily imply acceptable convergent validity or method effects. Therefore, item selection in order to improve convergent validity (or method influence) should be based on $\widehat{CO}$ (or $\widehat{MS}$) rather than on $\widehat{Rel}$. Although $\widehat{CO}$ can still be heightened to a required level for small $\widehat{\overline{\omega}}$ (the sample $\overline{\omega}$), it works only when $\widehat{\overline{\omega}} > \widehat{\overline{\psi}}$ (the sample $\overline{\psi}$). A required level of $\widehat{CO}$ may not be achieved if the items used are those with $\widehat{\overline{\psi}} \geq \widehat{\overline{\omega}}$. Therefore, item selection based on $\widehat{CO}$ should satisfy the requirement that $\widehat{\overline{\omega}} > \widehat{\overline{\psi}}$. When selecting items measured by nonstandard methods, we need those with larger $\widehat{\lambda}_{F_j}$ (the sample λ_{F_j}) and smaller $\widehat{\lambda}_{M_j}$ (the sample λ_{M_j}) to help achieve the goal of acceptable consistency coefficient, since their inclusion can lead to a higher $\widehat{\overline{\omega}}$ and a lower $\widehat{\overline{\psi}}$. Similarly, item selection based on $\widehat{MS}$ should satisfy the requirement that $\widehat{\overline{\psi}} > \widehat{\overline{\omega}}$. Eliminating an item with small $\widehat{\lambda}_{F_j}$ and large $\widehat{\lambda}_{M_j}$ may increase $\widehat{CO}$, and eliminating an item with small $\widehat{\lambda}_{M_j}$ and large $\widehat{\lambda}_{F_j}$ may increase $\widehat{MS}$, but the elimination may not increase $\widehat{Rel}$.

An alternative approach for scale reduction

If the number of items for a trait under MTMM designs is large, item elimination would be needed to achieve parsimony. One important reason to use short scales is the reduction of being fatigued or bored resulting from answering a lengthy set of items (Lindell & Whitney, [2001](#); Schmitt & Stults, [1985](#)). The properties of the consistency coefficient and the method-specificity coefficient discussed in the previous section can help assess the quality of scale items. An alternative approach for scale

reduction based on $\widehat{CO}$ (or $\widehat{MS}$) and the properties discovered is proposed. Since the summated score does not include the items measured by the standard method, only those measured by nonstandard methods are considered to be eliminated.

If the goal is to enhance convergent validity, we eliminate, one at a time during the process of scale reduction, the item so that the increment of $\widehat{CO}$ (denoted by $\Delta \widehat{CO}$) after item elimination is maximized. We stop the process when the maximum $\Delta \widehat{CO}$ becomes negative. In other words, no further elimination is allowed if convergent validity becomes worse. If the goal is to strengthen method influence, we sequentially eliminate the item for which the increment of $\widehat{MS}$ (denoted by $\Delta \widehat{MS}$) is maximized. We stop the elimination when the maximum $\Delta \widehat{MS}$ becomes negative. The specific step-by-step procedure under the purpose of maximizing convergent validity is given as follows:

> Step 0. Compute the $\widehat{CO}$ for the initial items measured by nonstandard methods.
> Step 1. Find the item whose elimination yields the maximum $\widehat{CO}$ from those left in the scale. If there is only one item associated with a nonstandard method, that item is excluded from those considered to be eliminated. Update the $\widehat{CO}$ and compute $\Delta \widehat{CO} =$ new $\widehat{CO}$ − old $\widehat{CO}$.
> Step 2. If $\Delta \widehat{CO} \geq 0$, then eliminate the item from the scale and then go to Step 1; otherwise, stop by returning the items left as the final choices, with the consistency coefficient of the most recently updated $\widehat{CO}$.

The procedure under the purpose of maximizing method effects is similar and need not be repeated. The final reduced scale consists of the items measured by the standard method and those determined according to the procedure given above.

It is required that no nonstandard method be neglected for measuring the items in the reduced scale. Item elimination is not allowed for any trait-method unit with only one item. The trait (method) loadings of the items in the reduced scale need to be significant and $\widehat{\overline{\omega}}$ needs to be greater (less) than $\widehat{\overline{\psi}}$ when the purpose is to maximize convergent validity (method influence). Moreover, discriminant validity must be maintained for the reduced scale achieving the maximum of convergent validity or method effects. If any of the previously mentioned requirements is not met for the reduced scale, scale reduction fails. The proposed approach, also applicable for subscales developed for sub-

constructs, is recommended for scale development based on MTMM designs.

Illustration

To illustrate the approach, we use a data set generated from the model depicted in Fig. 1, in which the first method was chosen as the standard method. x_{311} and x_{321} are measured by the standard method and will not be eliminated. x_{312}, x_{322}, x_{332}, x_{313}, and x_{323} are the items measured by nonstandard methods for F_3, which are to be further selected. The model parameters for data generation are given in Table 1 (in standard score form for ease of manipulation). They are determined based on Conway et al. (2004). The procedure given in Fan, Felsovalyi, Sivo, and Keenan (2002, Sections 4.3 and 7.2) is followed to generate a sample ($N = 300$) of normally distributed items, with which further item selection among x_{312}, x_{322}, x_{332}, x_{313}, and x_{323} by using the proposed approach is demonstrated. The sample correlation matrix of the items is given in Table 2.

The parameter estimates (obtained by using SAS PROC CALIS [SAS Institute Inc., 2010]) during the scale reduction process based on the CT-C($M-1$) model with M_1 as the standard method under the purpose of maximizing convergent validity or method influence are summarized in Table 3. At each step, $\widehat{\overline{\omega}}_{F_{3\_1}}$, $\widehat{\overline{\psi}}_{F_{3\_1}}$, $\widehat{\Upsilon}_{F_{3\_1}}$, $\widehat{Rel}_{F_{3\_1}}$, and $\widehat{CO}_{F_{3\_1}}$ (or $\widehat{MS}_{F_{3\_1}}$) are reported for the items retained in the scale, and $\widehat{\lambda}_{F_{3jk}}$ and $\widehat{\lambda}_{M_{3jk}}$ are given for the item removed. It appears that model fit is adequate during the scale reduction process and

Table 1 Population parameters of the CT-C($M-1$) model in Fig. 1 (for data generation)

Trait	Item	$\lambda_{F_{ijk}}$	$\lambda_{M_{ijk}}$	$\overline{\omega}_{F_{i\_1}}$	$\overline{\psi}_{F_{i\_1}}$	$CO_{F_{i\_1}}$	$MS_{F_{i\_1}}$	$Rel_{F_{i\_1}}$
F_{2_1}	x_{211}	.80		.688	.169	.741	.182	.923
	x_{212}	.75	.60					
	x_{213}	.60	.30					
F_{3_1}	x_{311}	.80		.183	.281	.320	.492	.812
	x_{321}	.70						
	x_{312}	.50	.60					
	x_{322}	.50	.50					
	x_{332}	.20	.55					
	x_{313}	.60	.40					
	x_{323}	.20	.80					
F_{4_1}	x_{411}	.70		.567	.131	.708	.164	.872
	x_{412}	.65	.40					
	x_{413}	.80	.30					
	x_{423}	.75	.50					

$\rho_{F_{i\_1}F_{i'\_1}} = .50 (i \neq i')), \rho_{M_2M_3} = .50$

parameter estimates are close to the corresponding population values, justifying our computational results.

In the first part (when maximizing $\widehat{CO}_{F_{3\_1}}$) of Table 3, $\widehat{\overline{\omega}}_{F_{3\_1}}(= .172) < \widehat{\overline{\psi}}_{F_{3\_1}}(= .278)$ at Step 0 for the five initial items, failing to meet the requirement and implying that $\widehat{CO}_{F_{3\_1}} < \widehat{CO}_{F_{3\_1},M} < .5$. It appears that some item may need to be removed. The item with the maximum increment of $\widehat{CO}_{F_{3\_1}}$ after its elimination is x_{323} ($\widehat{CO}_{F_{3\_1}}$ increases from .307 to .352 with $\Delta\widehat{CO}_{F_{3\_1}} = .045$). Therefore, x_{323} was first eliminated. However, $\widehat{\overline{\omega}}_{F_{3\_1}}(= .213)$ is still less than $\widehat{\overline{\psi}}_{F_{3\_1}}(= .259)$ after excluding x_{323}. Next, since x_{313} is the only item measured by M_3 among the rest of the four items, it was compulsorily included. x_{332}, with the largest $\Delta\widehat{CO}_{F_{3\_1}}(= .088)$ among the rest of the three items, was subsequently removed. $\widehat{\overline{\omega}}_{F_{3\_1}}(= .303)$, after excluding x_{323} and x_{332}, becomes greater than $\widehat{\overline{\psi}}_{F_{3\_1}}(= .229)$, indicating an improvement of the average item quality. The requirement of $\widehat{\overline{\omega}}_{F_{3\_1}} > \widehat{\overline{\psi}}_{F_{3\_1}}$ when maximizing convergent validity has also been met. Since the elimination of any item left would worsen convergent validity, the process stopped with the final retained items of x_{312}, x_{322}, and x_{313}, measured by nonstandard methods. $\widehat{CO}_{F_{3\_1}}$ improves from .307 to .44. Although $\widehat{Rel}_{F_{3\_1}}$ decreases from .804 down to .773, it should not receive concern since the purpose is to maximize convergent validity. At each step, estimates of the trait and method loadings are all highly significant ($p < .01$). Moreover, trait correlations show little change and all remain significantly less than 1 (p values associated with the chi-square difference tests < .001), implying that discriminant validity has been maintained while improving convergent validity by removing some "poor" items based on the consistency coefficient.

In the second part (when maximizing $\widehat{MS}_{F_{3\_1}}$) of Table 3, the final items selected according to a similar procedure include x_{312}, x_{332}, and x_{323}. x_{313} and x_{322} were sequentially removed on the basis of the criterion of achieving the largest $\Delta\widehat{MS}_{F_{3\_1}}$. For the five initial items, although $\widehat{\overline{\psi}}_{F_{3\_1}}(= .278) > \widehat{\overline{\omega}}_{F_{3\_1}}(= .172)$; hence, $\widehat{MS}_{F_{3\_1}} > \widehat{MS}_{F_{3\_1},M}$, $\widehat{\overline{\psi}}_{F_{3\_1}}$ is far less than .5, reflecting a need to make improvements. x_{313} was first eliminated because its elimination can lead to the maximum increase of $\widehat{MS}_{F_{3\_1}}$ $\left(\Delta\widehat{MS}_{F_{3\_1}} = .052\right)$. The new $\widehat{\overline{\psi}}_{F_{3\_1}}(= .331)$ becomes greater and the difference between $\widehat{\overline{\psi}}_{F_{3\_1}}$ and $\widehat{\overline{\omega}}_{F_{3\_1}}$ $(= .14)$ after elimination becomes bigger. The next one to be eliminated is selected from the rest of the four except x_{323} because x_{323} is the only item associated with M_3. x_{322} was removed because its elimination results in the

Table 2 Correlation matrix of the items resulting from a data set ($N = 300$) generated from the CT-C(M-1) model in Fig. 1 with parameters shown in Table 1

	x_{211}	x_{212}	x_{213}	x_{311}	x_{321}	x_{312}	x_{322}	x_{332}	x_{313}	x_{323}	x_{411}	x_{412}	x_{413}	x_{423}
x_{211}	1.000													
x_{212}	.629	1.000												
x_{213}	.529	.568	1.000											
x_{311}	.297	.314	.262	1.000										
x_{321}	.265	.237	.171	.610	1.000									
x_{312}	.172	.481	.260	.447	.362	1.000								
x_{322}	.195	.470	.230	.472	.398	.542	1.000							
x_{332}	.072	.361	.121	.141	.149	.481	.322	1.000						
x_{313}	.185	.300	.331	.495	.410	.408	.367	.230	1.000					
x_{323}	.050	.252	.320	.196	.077	.291	.221	.219	.419	1.000				
x_{411}	.239	.233	.179	.269	.212	.175	.284	.093	.134	.055	1.000			
x_{412}	.252	.468	.268	.288	.243	.494	.400	.314	.206	.204	.423	1.000		
x_{413}	.276	.342	.306	.350	.291	.285	.305	.128	.276	.286	.510	.561	1.000	
x_{423}	.262	.428	.357	.314	.263	.361	.340	.194	.438	.504	.435	.582	.681	1.000
M	−.098	−.031	−.109	−.074	−.004	.014	.001	.034	−.045	−.081	.037	.017	−.010	−.014
SD	.960	.968	1.015	1.090	1.080	1.017	.999	1.016	.988	.966	.963	.991	.948	.940

maximum $\Delta\widehat{MS}_{F_{3\_1}}(= .06)$. $\widehat{\overline{\psi}}_{F_{3\_1}}(= .405)$ becomes much greater than $\widehat{\overline{\omega}}_{F_{3\_1}}(= .093)$ after removing x_{313} and x_{322}. No further reduction was allowed. $\widehat{MS}_{F_{3\_1}}$ improves from .497 to .609 whereas $\widehat{Rel}_{F_{3\_1}}$ is decreasing. Again, at each step, trait loadings and method loadings are all significant, and trait correlations show little change and are all significantly less than 1 (p values < .001). Method effects have been enhanced by item elimination while maintaining discriminant validity.

The items x_{311} and x_{321} measured by the standard method are included in the final reduced scale, regardless of the purpose of maximizing convergent validity or method influence. However, the selection of the items measured by the nonstandard methods depends on the research purpose. If the purpose is to maximize convergent validity, then x_{312}, x_{322}, x_{313} are the choices; if the purpose is to maximize method influence, then x_{312}, x_{332}, x_{323} are the choices. It appears that the items selected under different purposes are somewhat different.

Table 3 Parameter estimates during the process of scale reduction for F_3 under the purpose of maximizing convergent validity or maximizing method effects based on the CT-C($M-1$) model with the first method as the standard method

Step	Item Removed	Items Retained	$\widehat{\lambda}_{F_{3jk}}$	$\widehat{\lambda}_{M_{3jk}}$	$\widehat{\overline{\omega}}_{F_{3\_1}}$	$\widehat{\overline{\psi}}_{F_{3\_1}}$	$\widehat{Rel}_{F_{3\_1}}$	$\widehat{CO}_{F_{3\_1}}(\widehat{MS}_{F_{3\_1}})$	$\widehat{\Upsilon}_{F_{3\_1}}$	$\widehat{\rho}_{F_{2\_1}F_{3\_1}}$	$\widehat{\rho}_{F_{2\_1}F_{4\_1}}$	$\widehat{\rho}_{F_{3\_1}F_{4\_1}}$
Maximizing Convergent Validity												
0		$x_{312}, x_{322}, x_{332}, x_{313}, x_{323}$			.172	.278	.804	.307	.238	.448***#	.460***#	.504***#
1	x_{323}	$x_{312}, x_{322}, x_{332}, x_{313}$	.164**	.778***	.213	.259	.782	.352	.232	.446***#	.463***#	.496***#
2	x_{332}	$x_{312}, x_{322}, x_{313}$	.165**	.575***	.303	.229	.773	.440	.210	.446***#	.460***#	.497***#
Maximizing Method Effects												
0		$x_{312}, x_{322}, x_{332}, x_{313}, x_{323}$			.172	.278	.804	(.497)	.238	.448***#	.460***#	.504***#
1	x_{313}	$x_{312}, x_{322}, x_{332}, x_{323}$	.553***	.460***	.140	.331	.781	(.549)	.279	.452***#	.465***#	.523***#
2	x_{322}	$x_{312}, x_{332}, x_{323}$	.546***	.412***	.093	.405	.748	(.609)	.340	.444***#	.468***#	.510***#

Only the items measured by M_2 and M_3 (the nonstandard methods) are considered to be eliminated. $\widehat{\lambda}_{F_{3jk}}$ and $\widehat{\lambda}_{M_{3jk}}$ are the estimates associated with the item removed at each step. The model fit was adequate during the process (p values associated with the chi-square tests were all greater than .5)

** $p < .01$ *** $p < .001$

significantly less than 1

Discussion and conclusion

The interpretative and estimation problems that the CT–CM model suffers from can be overcome by the CT-C(M−1) model. Under the CT-C(M − 1) model, it has been shown that the reliability coefficient, the consistency coefficient, and the method-specificity coefficient of the sum of the items measured with nonstandard methods for a trait factor are all functions of the average consistency coefficient ($\overline{\omega}$), the average method-specificity coefficient ($\overline{\psi}$), and scale length. It is noteworthy that using many items for abstract trait factors may heighten convergent validity to a required level only when $\widehat{\overline{\omega}} > \widehat{\overline{\psi}}$, but cannot when $\widehat{\overline{\omega}} \leq \widehat{\overline{\psi}}$. In addition, given a fixed level of convergent validity, the number of items used can be reduced by increasing $\overline{\omega}$ and further by decreasing $\overline{\psi}$. A general strategy is to eliminate the items with smaller trait loadings and larger method loadings, or equivalently, to take those with larger trait loadings and smaller method loadings. If the goal is to enhance method effects, then the items with larger method loadings and smaller trait loadings should be retained.

The same principles that we have shown to find properties for CO and MS based on the CT-C(M−1) model could also be applied to the CT–CM model.

An alternative approach to reduce scale length for MTMM data based on the CT-C(M−1) model under the goal of enhancing convergent validity or method influence has been proposed. Traditional approaches for scale reduction without taking into account method factors are inappropriate for MTMM data. The items eliminated during the reduction process based on commonly used criteria in practice such as "the lowest item-total correlation" and "the largest increment of coefficient alpha" could be quite different from those determined on the basis of the criterion of the largest non-negative increment of $\widehat{CO}$ (or $\widehat{MS}$) proposed in this study.

If the consistency coefficient resulting from the final reduced scale is still not large enough, then the criterion of convergent validity might be released to make a compromise. If this is not possible, then it becomes necessary to return to the initial stage of scale development and reconduct item generation and item analysis. A poor item for a trait factor should be replaced by another one that can reflect the same facet of the trait domain and that can lead to a higher $\widehat{\lambda}_F$ and a lower $\widehat{\lambda}_M$. In the meantime, the maintenance of discriminant validity is required. Researchers should make efforts to modify existing items and/or substitute new items to achieve acceptable convergent validity, discriminant validity, and parsimony. The suggestions regarding the method-specificity coefficient are similar. Once the final scale obtained is satisfactory, it needs to be further validated by using another independent sample, as usually seen in the literature.

Note that the items retained in the final reduced scale may depend on the standard method chosen in the CT-C(M−1) model, and the items measured by the standard method cannot be assessed because the trait factor is confounded with the standard method. It is therefore important that the choice of the standard method and the items measured by that method should be justified on the basis of theoretical considerations or consensus in the field. Following the recommendation given by Eid et al. (2003) that a "gold standard method" be chosen, we suggest that "reference items" (usually key items for the trait factor) be measured by that "gold standard method" and that other potential items be measured by nonstandard methods since item elimination is conducted only for those measured by nonstandard methods.

The multiple-item CT-C(M−1) model, in which there are multiple items for each trait-method unit, has been given to address the problem of trait-specific method effects (Eid et al., 2003). Analyses of CO, MS, and Rel discussed in the present study can be extended for the multiple-item CT-C(M−1) model. Statistical inference for reliability-related parameters in the model, particularly for CO and MS, needs to be further studied. In addition, how the effectiveness of the proposed approach for scale reduction is influenced by sample size is an interesting task for future research.

Appendix

Derivations of Eqs. 5, 6, and 7

By choosing the first method ($k = 1$) as the standard method for the CT-C(M−1) model, the variance of the summated score $\sum\limits_{k=2}^{K} \sum\limits_{j=1}^{J_{ik}} x_{ijk}$ under the assumptions for Eq. 4 and by fixing $\sigma^2_{F_{i\_1}}$ and $\sigma^2_{M_k}$ to 1 for identification reasons is given by

$$Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}x_{ijk}\right) = Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}t_{ijk}\right) + Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\delta_{ijk}\right)$$

$$= Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\lambda_{F_{ijk}}F_{i\_1}\right) + Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\lambda_{M_{ijk}}M_k\right) + Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\delta_{ijk}\right)$$

$$= \left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\lambda_{F_{ijk}}\right)^2 + \sum_{k=2}^{K}\left(\sum_{j=1}^{J_{ik}}\lambda_{M_{ijk}}\right)^2 + \sum_{k\neq k'}\left(\sum_{j=1}^{J_{ik}}\lambda_{M_{ijk}}\right)\left(\sum_{j=1}^{J_{ik'}}\lambda_{M_{ijk'}}\right)\rho_{M_kM_{k'}} + \sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\sigma^2_{\delta_{ijk}}$$

$$= J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \sum_{k=2}^{K}J_{ik}^2\left(\overline{\lambda}_{M_{ik}}\right)^2 + \sum_{k\neq k'}\left(J_{ik}\overline{\lambda}_{M_{ik}} \times J_{ik'}\overline{\lambda}_{M_{ik'}}\right)\rho_{M_kM_{k'}} + J_i\overline{\sigma^2_{\delta_i}}$$

$$= J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + J_i^2\sum_{k=2}^{K}\left(w_k\overline{\lambda}_{M_{ik}}\right)^2 + J_i^2\sum_{k\neq k'}\left(w_k\overline{\lambda}_{M_{ik}}\right)\left(w_{k'}\overline{\lambda}_{M_{ik'}}\right)\rho_{M_kM_{k'}} + J_i\overline{\sigma^2_{\delta_i}}$$

$$= J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + J_i^2\Upsilon_{F_{i\_1}} + J_i\overline{\sigma^2_{\delta_i}},$$

where $J_i = \sum_{k=2}^{K}J_{ik}$, $\overline{\lambda}_{F_{i\_1}} = \sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\lambda_{F_{ijk}}/J_i$ (the mean trait loading), $\overline{\lambda}_{M_{ik}} = \sum_{j=1}^{J_{ik}}\lambda_{M_{ijk}}/J_{ik}$ (the mean method loading for M_k), $\Upsilon_{F_{i\_1}} = \sum_{k=2}^{K}\left(w_k\overline{\lambda}_{M_{ik}}\right)^2 + \sum_{k\neq k'}\left(w_k\overline{\lambda}_{M_{ik}}\right)\left(w_{k'}\overline{\lambda}_{M_{ik'}}\right)\rho_{M_kM_{k'}}$, $\overline{\sigma^2_{\delta_i}} = \sum_{k=1}^{K}\sum_{j=1}^{J_{ik}}\sigma^2_{\delta_{ijk}}/J_i$, and $w_k = J_{ik}/J_i$ $(k > 1)$. w_k represents the weight for the kth method based on the number of items associated with that method, $\sum_{k=2}^{K}\left(w_k\overline{\lambda}_{M_{ik}}\right)^2$ is the sum of squares of $\overline{\lambda}_{M_{ik}}$ weighted by w_k, and $\sum_{k\neq k'}\left(w_k\overline{\lambda}_{M_{ik}}\right)\left(w_{k'}\overline{\lambda}_{M_{ik'}}\right)\rho_{M_kM_{k'}}$ is the sum of products of the weighted $\overline{\lambda}_{M_{ik}}$, the weighted $\overline{\lambda}_{M_{ik'}}$, and their method correlations. It follows that the population reliability coefficient $Rel_{F_{i\_1}}$ of $\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}x_{ijk}$ with respect to F_{i_1} is given by

$$Rel_{F_{i\_1}} = \frac{Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}t_{ijk}\right)}{Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}x_{ijk}\right)}$$

$$= \frac{J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + J_i^2\Upsilon_{F_{i\_1}}}{J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + J_i^2\Upsilon_{F_{i\_1}} + J_i\overline{\sigma^2_{\delta_i}}}$$

$$= \frac{J_i\left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}}\right]}{(J_i - 1)\left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}}\right] + \left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma^2_{\delta_i}}\right]}$$

$$= \frac{J_i\overline{\theta}^*_{F_{i\_1}}}{1 + (J_i - 1)\overline{\theta}^*_{F_{i\_1}}},$$

where $\overline{\theta}^*_{F_{i\_1}} = \frac{\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}}}{\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma^2_{\delta_i}}}$. Moreover, the consistency coefficient $CO_{F_{i\_1}}$ and the method-specificity coefficient $MS_{F_{i\_1}}$ of $\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}x_{ijk}$ with respect to F_{i_1} are given by

$$CO_{F_{i\_1}} = \frac{Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\lambda_{F_{ijk}}F_{i\_1}\right)}{Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}x_{ijk}\right)}$$

$$= \frac{J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2}{J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + J_i^2\Upsilon_{F_{i\_1}} + J_i\overline{\sigma^2_{\delta_i}}}$$

$$= \frac{J_i\left(\overline{\lambda}_{F_{i\_1}}\right)^2}{(J_i - 1)\left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}}\right] + \left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma^2_{\delta_i}}\right]}$$

$$= \frac{J_i\overline{\omega}_{F_{i\_1}}}{1 + (J_i - 1)\overline{\theta}^*_{F_{i\_1}}},$$

and

$$MS_{F_{i\_1}} = \frac{Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}\lambda_{M_{ijk}}M_k\right)}{Var\left(\sum_{k=2}^{K}\sum_{j=1}^{J_{ik}}x_{ijk}\right)}$$

$$= \frac{J_i^2\Upsilon_{F_{i\_1}}}{J_i^2\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + J_i^2\Upsilon_{F_{i\_1}} + J_i\overline{\sigma^2_{\delta_i}}}$$

$$= \frac{J_i\Upsilon_{F_{i\_1}}}{(J_i - 1)\left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}}\right] + \left[\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma^2_{\delta_i}}\right]}$$

$$= \frac{J_i\overline{\psi}_{F_{i\_1}}}{1 + (J_i - 1)\overline{\theta}^*_{F_{i\_1}}},$$

where $\overline{\omega}_{F_{i\_1}} = \dfrac{\left(\overline{\lambda}_{F_{i\_1}}\right)^2}{\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma^2_{\delta_i}}}$ and $\overline{\psi}_{F_i} = \dfrac{\Upsilon_{F_{i\_1}}}{\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \Upsilon_{F_{i\_1}} + \overline{\sigma^2_{\delta_i}}}$. $\overline{\theta}^*_{F_{i\_1}} = \overline{\omega}_{F_{i\_1}} + \overline{\psi}_{F_{i\_1}}$. $\overline{\omega}_{F_{i\_1}}$ and $\overline{\psi}_{F_{i\_1}}$ indicate, respectively, the average proportion of variance due to F_{i_1} and the average proportion of variance due to nonstandard methods. For the special case in which only one nonstandard method is used, we have $\Upsilon_{F_{i\_1}} = \left(\overline{\lambda}_{M_{i2}}\right)^2$,

$\overline{\omega}_{F_{i\_1}} = \dfrac{\left(\overline{\lambda}_{F_{i\_1}}\right)^2}{\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \left(\overline{\lambda}_{M_{i2}}\right)^2 + \overline{\sigma^2_{\delta_i}}}$, and $\overline{\psi}_{F_{i\_1}} = \dfrac{\left(\overline{\lambda}_{M_{i2}}\right)^2}{\left(\overline{\lambda}_{F_{i\_1}}\right)^2 + \left(\overline{\lambda}_{M_{i2}}\right)^2 + \overline{\sigma^2_{\delta_i}}}$.

Denoting the sample $CO_{F_{i\_1}}$, $\overline{\omega}_{F_{i\_1}}$, and $\overline{\psi}_{F_{i\_1}}$ by $\widehat{CO}_{F_{i\_1}}$, $\widehat{\overline{\omega}}_{F_{i\_1}}$, and $\widehat{\overline{\psi}}_{F_{i\_1}}$, we have, by following Eq. 6, $\widehat{CO}_{F_{i\_1}} =$

$\dfrac{J_i \widehat{\overline{\omega}}_{F_{i\_1}}}{1 + (J_i - 1)\left(\widehat{\overline{\omega}}_{F_{i\_1}} + \widehat{\overline{\psi}}_{F_{i\_1}}\right)}$, $\widehat{\overline{\omega}}_{F_{i\_1}} = \dfrac{\left(\widehat{\overline{\lambda}}_{F_{i\_1}}\right)^2}{\left(\widehat{\overline{\lambda}}_{F_{i\_1}}\right)^2 + \widehat{\Upsilon}_{F_{i\_1}} + \widehat{\overline{\sigma^2_{\delta_i}}}}$, $\widehat{\overline{\psi}}_{F_{i\_1}} =$

$\dfrac{\widehat{\Upsilon}_{F_{i\_1}}}{\left(\widehat{\overline{\lambda}}_{F_{i\_1}}\right)^2 + \widehat{\Upsilon}_{F_{i\_1}} + \widehat{\overline{\sigma^2_{\delta_i}}}}$, where $\widehat{\Upsilon}_{F_{i\_1}} = \sum\limits_{k=2}^{K}\left(w_k \widehat{\overline{\lambda}}_{M_{ik}}\right)^2 + \sum\limits_{k \neq k'}$

$\left(w_k \widehat{\overline{\lambda}}_{M_{ik}}\right)\left(w_{k'} \widehat{\overline{\lambda}}_{M_{ik'}}\right)\widehat{\rho}_{M_k M_{k'}}$, $\widehat{\overline{\lambda}}_{F_{i\_1}} = \sum\limits_{k=2}^{K}\sum\limits_{j=1}^{J_{ik}}\widehat{\lambda}_{F_{ijk}}/J_i$, $\widehat{\overline{\lambda}}_{M_{ik}} =$

$\sum\limits_{j=1}^{J_{ik}}\widehat{\lambda}_{M_{ijk}}/J_{ik}$, $J_i = \sum\limits_{k=2}^{K}J_{ik}$, and $\widehat{\overline{\sigma^2_{\delta_i}}} = \sum\limits_{k=2}^{K}\sum\limits_{j=1}^{J_{ik}}\widehat{\sigma}^2_{\delta_{ijk}}/J_i$. In addi-

tion, $\widehat{CO}_{F_{i\_1},U} = \dfrac{J_i \widehat{\overline{\omega}}_{F_{i\_1}}}{1 + (J_i - 1)\widehat{\overline{\omega}}_{F_{i\_1}}}$, $\widehat{CO}_{F_{i\_1},M} = \dfrac{J_i \widehat{\overline{\omega}}_{F_{i\_1}}}{1 + 2(J_i - 1)\widehat{\overline{\omega}}_{F_{i\_1}}}$, and

$\widehat{CO}_{F_{i\_1},L} = \widehat{\overline{\omega}}_{F_{i\_1}}$. Similarly, by following Eq. 7, we have

$\widehat{MS}_{F_{i\_1}} = \dfrac{J_i \widehat{\overline{\psi}}_{F_{i\_1}}}{1 + (J_i - 1)\left(\widehat{\overline{\omega}}_{F_{i\_1}} + \widehat{\overline{\psi}}_{F_{i\_1}}\right)}$. Moreover, $\widehat{MS}_{F_{i\_1},U} =$

$\dfrac{J_i \widehat{\overline{\psi}}_{F_{i\_1}}}{1 + (J_i - 1)\widehat{\overline{\psi}}_{F_{i\_1}}}$, $\widehat{MS}_{F_{i\_1},M} = \dfrac{J_i \widehat{\overline{\psi}}_{F_{i\_1}}}{1 + 2(J_i - 1)\widehat{\overline{\psi}}_{F_{i\_1}}}$, and $\widehat{MS}_{F_{i\_1},L} = \widehat{\overline{\psi}}_{F_{i\_1}}$.

References

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44,* 175–184.

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

Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling, 11,* 535–559.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 96–104.

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24,* 315–318.

Deinzer, R., Steyer, R., Eid, M., Kotz, P., Schwenkmezger, P., Ostendorf, F., & Neubauer, A. (1995). Situational effects in trait assessment: The FPI, NEOFFI, and EPI questionnaires. *European Journal of Personality, 9,* 1–23.

Eid, M. (2000). A multitrait–multimethod model with minimal assumptions. *Psychometrika, 65,* 241–261.

Eid, M., Lischetzke, T., Nussbeck, F., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait–multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8,* 38–60.

Eid, M., & Diener, E. (2006). Introduction: The need for multimethod measurement in psychology. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 3–8). Washington, DC: American Psychological Association.

Fan, X., Felsovalyi, A., Sivo, S. A., & Keenan, S. C. (2002). *SAS for Monte Carlo studies: A guide for quantitative researchers.* Cary, NC: SAS Institute Inc.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18,* 39–50.

Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M−1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods, 13,* 49–57.

Kenny, D. A., & Kashy, D. A. (1992). The analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165–172.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait–correlated method and correlated uniqueness models for multitrait–multimethod data. *Psychological Methods, 7,* 228–244.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86,* 114–121.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science, 52,* 1865–1883.

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., & Grayson, D. (1995). Latent variable models of multitrait–multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks, CA: Sage.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

Mishra, D. P. (2000). An empirical assessment of measurement error in health-care survey research. *Journal of Business Research, 48,* 193–205.

Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology, 37,* 687–702.

Pohl, S., Steyer, R., & Kraus, K. (2008). Modeling method effects as individual causal effects. *Journal of the Royal Statistical Society: Series A, 171,* 1–23.

SAS Institute Inc. (2010). *SAS/STAT user's guide* (SAS 9.22). Cary, NC: SAS Institute Inc.

Schmitt, M. J., & Steyer, R. (1993). A latent state-trait model (not only) for social desirability. *Personality and Individual Differences, 14,* 519–529.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9,* 367–373.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait- multimethod matrices. *Applied Psychological Measurement, 10,* 1–22.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait- multimethod data. *Applied Psychological Measurement, 9,* 1–26.