Published in final edited form as:

Struct Equ Modeling. 2014 ; 21(4): 509–523. doi:10.1080/10705511.2014.919816.

Data-Generating Mechanisms Versus Constructively-Defined Latent Variables in Multitrait-Multimethod Analysis: A Comment on Castro-Schilo, Widaman, and Grimm (2013)

Christian Geiser

Department of Psychology, Utah State University.

Tobias Koch and Michael Eid

Department of Education and Psychology, Freie Universität Berlin, Germany.

Abstract

In a recent article, Castro-Schilo, Widaman, and Grimm (2013) compared different approaches for relating multitrait-multimethod (MTMM) data to external variables. Castro-Schilo et al. reported that estimated associations with external variables were in part biased when either the Correlated Traits-Correlated Uniqueness (CT-CU) or Correlated Traits-Correlated (Methods – 1) [CT-C(M – 1)] models were fit to data generated from the Correlated Traits-Correlated Methods (CT-CM) model, whereas the data-generating CT-CM model accurately reproduced these associations. Castro-Schilo et al. argued that the CT-CM model adequately represents the data-generating mechanism in MTMM studies, whereas the CT-CU and CT-C(M – 1) models do not fully represent the MTMM structure. In this comment, we question whether the CT-CM model is more plausible as a data-generating model for MTMM data than the CT-C(M – 1) model. We show that the CT-C(M – 1) model can be formulated as a reparameterization of a basic MTMM true score model that leads to a meaningful and parsimonious representation of MTMM data. We advocate the use CFA-MTMM models in which latent trait, method, and error variables are explicitly and constructively defined based on psychometric theory.

Keywords

Multitrait-multimethod (MTMM) analysis; CT-CM model; CT-C(M – 1) model; constructively-defined latent variables; psychometric theory

The question as to how multitrait-multimethod (MTMM) data is most appropriately analyzed statistically has puzzled methodological and substantive researchers for decades—and continues to puzzle investigators in the present. In a recent paper, Castro-Schilo, Widaman, and Grimm (2013) examined an important question: How do different confirmatory factor analysis (CFA) models for analyzing MTMM data perform when it comes to studying the associations of trait and method influences with external variables? This question is an important one for multiple reasons. MTMM researchers are often

interested in relating trait and method factors to external variables, because this can help us better understand the antecedents and consequences of trait and method effects. Furthermore, as Castro-Schilo et al. note, different CFA-MTMM models use different specifications of trait and method factors. Researchers may be unsure about how to properly interpret (1) the parameter estimates from different approaches and (2) differences in how different approaches represent relationships between trait and method factors and external variables. In the present paper, we comment on findings and conclusions presented in the Castro-Schilo et al. (2013) article.

Summary of the Castro-Schilo et al. (2013) Study

In their study, Castro-Schilo et al. examined the performance of different approaches for analyzing MTMM data with covariates (external variables used as predictors or outcomes of trait and method factors) using a simulation study (their Study 1) and an application to an actual data set (their Study 2). In their Study 1, Castro-Schilo et al. generated data based on the Correlated Traits-Correlated Methods (CT-CM) model and then fit the properly specified data-generating CT-CM model, the Correlated Traits-Correlated Uniqueness (CT-CU) and Correlated Traits-Correlated(Methods – 1) [CTC(M – 1)] models, as well as other models to the data. In the results section of their Study 1, Castro-Schilo et al. reported bias in the estimated associations of trait and/or method factors with external variables in the CT-CU and CT-C(M – 1) models as well as other approaches, whereas little bias was found for converged solutions of the data-generating CT-CM model. In the results section of their Study 2, Castro-Schilo et al. reported that estimated associations with external variables for the CTC(M – 1) and CT-CU models partly resulted in different conclusions about the magnitude and direction of relationships relative to the population CT-CM model.

One of the conclusions in the Castro-Schilo et al. article was that “MTMM structural models that do not specify all MVs [measured variables] as decomposable trait-method units [e.g., CT-CU and CT-C(M – 1)] partially ignore the true structure of the data” (p. 204) and that “fitting MTMM CFA models that are not true to the structure of the data (, ...) does not seem to be an optimal approach.” (p. 205). The claims made by Castro-Schilo et al. can thus be read as suggesting that users should prefer the CT-CM model over other approaches [such as the CT-CU and CT-C(M – 1) models], because only the CT-CM model provides a complete decomposition of each measured variable into trait, method, and error components, which, according to Castro-Schilo et al. is in line with the “true” MTMM structure (i.e., each measured variable is supposed to contain separable trait, method, and error components).

Reasons for the Present Comment

Our motivation for the present comment is threefold: First, we think that Castro-Schilo et al.’s use of the term “bias” in the context of comparisons between the CT-CM and other CFA-MTMM models may confuse both methodological and substantive researchers and that it requires clarification—particularly with respect to the comparison of results from the CT-CM and CT-C(M – 1) models. Second, we feel that Castro-Schilo et al.’s assumptions according to which (1) the CT-CM model represents the most plausible data-generating model for MTMM data in general and (2) the CTC(M – 1) model neglects part of the

MTMM data structure can be questioned, and that it is necessary to provide an additional, alternative perspective on this issue. Third, we want to lay out our arguments for why we think that the CT-C(M – 1) approach is in fact fully in line with Campbell and Fiske's (1959) MTMM matrix and show that it provides clear and unbiased results if the results are properly interpreted. We demonstrate that even though the CT-C(M – 1) approach does not specify trait and method factors for each trait-method unit (TMU), it still provides a complete and meaningful decomposition of MTMM data that avoids overparametrizations as encountered in the CT-CM model.

Overview

Our commentary is structured as follows: We first discuss what we see as problematic in the design, implementation, and interpretation of Castro-Schilo et al.'s simulation study and the line of argument put forth by these authors. We then show that Castro-Schilo et al.'s claim according to which the CT-CM model is the most attractive candidate for the data-generating mechanism in MTMM studies may be questioned based on past and present theoretical and empirical evidence against this model. Subsequently, we review Marsh and Hocevar's (1988) basic MTMM measurement model that allows conceptualizing each TMU in terms of well-defined latent true score variables from Classical Test Theory (CTT). We argue that this model is fully in line with Campbell and Fiske's (1959) MTMM approach and show that the CT-C(M – 1) model can be derived as a straightforward reparameterization of this model. We conclude that the CT-C(M – 1) model may be more plausible for MTMM data than the CT-CM model, because the CT-C(M – 1) model avoids an overparameterization of the basic latent MTMM structure. We clarify how findings from the CT-C(M – 1) approach are properly interpreted. In our discussion, we highlight the general advantages of CFA-MTMM models that are based on psychometric theories and that allow defining "trait", "method", and "error" latent variables clearly on the basis of well-defined true score variables. Finally, we provide arguments for why we think that our view does not contradict, but instead is fully in line with, Campbell and Fiske's (1959; Fiske & Campbell, 1992) MTMM approach.

Comment on the Castro-Schilo et al. (2013) Study

We begin our comment by making clear on which points we agree with Castro-Schilo et al. (2013). We fully agree with the authors that MTMM structures should not be ignored, but properly modeled, preferably with latent variable statistical models that account for random errors of measurement (i.e., appropriate CFA-MTMM models). We also agree with Castro-Schilo et al. that researchers often fail to acknowledge the fact that each measured variable represents a TMU (Campbell & Fiske, 1959) and that the issue of method-specificity, particularly in studies that use only a single-method design (e.g., self-report measures only), should not be ignored. We furthermore agree that different CFA-MTMM approaches have different strengths and weaknesses, and that different models can yield different parameter estimates that have different interpretations. Similar to Castro-Schilo et al., we think that it is important to study these differences and to clarify them for methodological and substantive researchers.

In their simulation, Castro-Schilo et al. generated data based on the CT-CM model as the “true” population model and then fit the (correctly-specified) CT-CM model as well as other models [CT-CU, CT-C(M – 1), and others] to the generated data. We do not think that Castro-Schilo et al.’s simulation approach is problematic per se. Simulating MTMM models can provide useful insights into the way different models work and/or under which conditions they show problems. We even think that it can be useful under certain circumstances to study what happens if a certain model (say, Model A) generated the data, but a different Model B is used to analyze it, as was done in Castro-Schilo et al. However, we think that (1) Castro-Schilo et al.’s interpretation of the results obtained from models other than the data-generating CT-CM model as “biased” is problematic and that (2) the argument that the CT-CM model is the only model among the three that completely represents the MTMM structure is questionable.

Interpretation of CT-CU and CT-C(M – 1) Parameter Estimates as “Biased”

In their Discussion, Castro-Schilo et al. stated that “...to the degree that researchers adhere to Campbell and Fiske’s (1959) conceptualization of variables as trait-method units, knowing the degree of bias that the CT-CU and CT-C(M – 1) models produce is of vital importance.” (p. 204). In this section, we explain why Castro-Schilo et al.’s interpretation of differences in parameter estimates obtained from different CFA-MTMM models as “bias” seems problematic to us. Further below, we show that the CT-C(M – 1) model is in fact fully in line with Campbell and Fiske’s (1959) concept of TMUs.

It seems little surprising that MTMM models different in structure from the model used to generate the data would in general have difficulties exactly reproducing the population parameters of the data-generating (in this case CT-CM) model, even if they imply a similar covariance structure. Why should we expect to obtain the same parameter estimates if the models that we fit are not the same as the model that generated the data? To illustrate, would we expect to obtain the same parameter estimates if we fit, for example, a five-factor CFA model to data that was generated based on a six factor CFA model?

This may seem like an unsuitable example at first sight. In fact, however, this is what happens when the CT-C(M – 1) model is fit to data generated based on the CT-CM model for a 3 traits $\times$ 3 methods MTMM design: The CT-CM model uses six (three trait and three method) factors to account for the data, whereas the CT-C(M – 1) model uses only five (three trait and two method) factors. Furthermore, as Castro-Schilo et al. acknowledge in their paper, trait and method factors are defined differently in the CT-C(M – 1) and CT-CM models. If these factors do not have the same theoretical meaning, why should we expect them to correlate in the same way with external variables? We elaborate on this point later in this comment when we discuss the CT-C(M – 1) model in detail.

In our view, Castro-Schilo et al.’s approach provides relatively little insight into the soundness (or lack of soundness) of the CT-CU and CT-C(M – 1) models to MTMM analyses. It does provide some insights into the lack of soundness of the CT-CM model, however, as we show later on.

Selecting one model as population model and fitting other models to data generated from this model, finding discrepancies, and concluding that the alternative models yield biased results seems almost like a circular argument to us. Consider another example: if we fit the Rasch (1-parameter logistic) model to item response data generated from a Birnbaum (2-parameter logistic) model, it is likely that we will find that the Rasch model produces “biased” estimates of, for example, correlations of the estimated latent trait scores with external variables relative to the population (Birnbaum) model. Can we thus conclude that the Rasch model is less suitable in general to fit item response data than the Birnbaum model?

In order for the line of reasoning presented in Castro-Schilo et al. to hold, we would need to assume that we can be *certain* about the fact that the CT-CM model is the only model that qualifies as the data-generating model in MTMM studies in general. We argue here that even though Castro-Schilo et al. seemed to make this assumption more or less explicitly in their article as shown below, the argument (1) does not appear to receive support from either their simulation study or empirical application and (2) seems questionable in general based on the large body of theoretical, statistical, and empirical evidence against the soundness of the CT-CM model. In the following section, we first review the CT-CM model and summarize previous critiques of this model. Subsequently, we present the CT-C(M – 1) model and explain why we think that this model may be more plausible and useful for analyzing MTMM data than the CT-CM model.

The CT-CM Model

The CT-CM model uses as many trait factors as there are distinct constructs in an MTMM study and as many method factors as there are different methods. In the model, all T trait factors are allowed to correlate, and all M method factors are allowed to correlate. Correlations between trait and method factors are not allowed.

In their article, Castro-Schilo argued that “In the first [simulation] study [Study 1], we generated MTMM data with characteristics that are likely to exist in empirical data sets.” (p. 204) and that “the CT-CM model maps directly onto the conceptualization of MTMM data put forth by Campbell and Fiske (1959), which is why it is so attractive theoretically.” (p. 206). Thus, we may speculate that Castro-Schilo et al.’s rationale for choosing the CT-CM model as population model is as follows: Generating data from the CT-CM model (rather than any of the available alternatives) makes most sense, because the CT-CM model is the model that is most closely in line with Campbell and Fiske’s original idea of the MTMM matrix and trait-method units (TMUs). That is, in the MTMM matrix *each* measured variable represents a TMU, and therefore, trait and method effects should be separable for each variable.

In contrast to the CT-CM model, the CT-CU and CT-C(M – 1) models do not allow for such a decomposition, because (1) the CT-CU model specifies no method factors and (2) the CT-C(M – 1) model uses a reference method for which no method factor is included, and contrasts this reference method against the other methods. The CT-CU and CT-C(M – 1) models therefore seem less suitable as population (“data-generating”) models in Castro-

Schilo et al.'s view, because these models do not allow decomposing each measured variable into "trait", "method", and "error" components.

Can we be sure that the CT-CM model is necessarily a better candidate for the "data-generating model" in MTMM studies than other models? We argue that this is not necessarily the case, and find support for our skepticism in both the prior theoretical and empirical literature examining the properties of the CT-CM model and in results of Castro-Schilo et al.'s simulation study and actual data application.

It is well-known from the prior literature cited in Castro-Schilo et al. that the CT-CM model is problematic in terms of model identification, convergence, improper solutions (e.g., negative variance estimates), and interpretation difficulties of resulting parameter estimates (Eid, 2000; Kenny & Kashy, 1992; Marsh, 1989). Perhaps most importantly, the CT-CM model is not globally identified (Grayson & Marsh, 1994). That is, the model is identified only for some, but not all possible constellations of parameters (Kenny & Kashy, 1992; Steyer, 1995).

For example, the CT-CM model is not identified when all trait loadings are equal and all method loadings are equal in the population, a condition that would actually seem desirable and parsimonious in practice. Castro-Schilo et al. dealt with this issue by specifying larger trait loadings for the first method in their simulation design than for the remaining methods. We think that there is, however, no good theoretical reason why one method should always differ from all other methods in terms of the loading pattern. More importantly, this requirement leads to the undesirable restriction that in practice, methods "have to" differ in their loading pattern for the model to work.

The convergence and improper solution problems observed by Castro-Schilo et al. seem to occur in the majority of applications of the CT-CM model to either empirical or simulated data (e.g., Marsh, 1989). Furthermore, Marsh (1989) discussed conceptual issues with the CT-CM model that lead to interpretation problems when all trait and all method factors are correlated.

Castro-Schilo et al. themselves acknowledged that "[the CT-CM] model resulted in many convergence issues, many out-of-bounds [i.e., improper parameter] estimates, and large variability in the estimates across data sets in the condition where external variables were predictors of trait and method factors." (pp. 197-198). In Castro-Schilo et al.'s simulations, the CT-CM model resulted in up to 69% non-converged replications, even though the CT-CM model used to analyze the simulated data was correctly specified (i.e., identical to the data-generating model). In addition, Castro-Schilo et al. reported "many out-of bounds [i.e., improper] estimates" for the CT-CM model (p. 197). That is, of the converged CT-CM solutions, apparently many resulted in improper parameter estimates, even though the very same model generated the data and the simulated sample size was substantial ($N = 500$).

Castro-Schilo et al. did not report the exact percentage of improper solutions for the converged CT-CM solutions for any of their simulation conditions, but from prior simulation studies, we know that the percentages are high (e.g., Marsh, 1989; Marsh &

Bailey, 1991). In addition, Castro-Schilo et al. obtained an improper solution in their application of the CT-CM model to actual data (their Study 2).

Castro-Schilo et al. mentioned that in their simulation, non-converged replications were not replaced by new converged replications and that the reported simulation results were based on only the remaining converged solutions. This means that in the most extreme case, results reported in Castro-Schilo et al. were based on only 31 data sets out of the total number of 100 replications. In our view, the low number of available solutions for some of the conditions causes additional problems. How confident can we be that the results provide a reliable picture of the performance of the CT-CM model, given the low number of converged solutions in several of the conditions?

Therefore, Castro-Schilo et al.'s claim according to which "The simulation study suggests that the CT-CM model does the best job at recapturing the population parameters of the associations of external variables with trait and method factors" (p. 204) seems premature to us. Instead, it appears to us that the Castro-Schilo study replicates findings from the previous literature highlighting the severe theoretical and empirical problems with the CT-CM model. It seems questionable to us whether a model that is not globally identified and for which even with simulated data in correctly specified and theoretically identified conditions, estimation problems are abundant, can be viewed as a theoretically attractive candidate for the "true" data-generating mechanism for MTMM structures in general or for modeling MTMM data in empirical applications.

From our perspective, the CT-CM model represents an overparametrized structure for MTMM data, thus creating the typical problems of empirical underidentification, non-convergence, and improper solutions often seen for overfitted SEM models (e.g., Rindskopf, 1984). In the following section, we explain why we think that the CT-C(M – 1) model may in fact be a more plausible model for MTMM data than the CT-CM model. We show that in contrast to the CT-CM model, the CT-C(M – 1) model does not overfit MTMM structures. We first review the basic MTMM measurement model proposed by Marsh and Hocevar (1988) and show that this model can be reparameterized as a CT-C(M – 1) model. To make our argument clear, we refer to MTMM designs with multiple as opposed to only single indicators per TMU. MTMM designs with a single indicator per TMU require the very strong assumption that method effects are perfectly correlated across different constructs to obtain identified CFA-MTMM models. This assumption is very restrictive and typically unrealistic for empirical applications. For example, the method effect of a Method A, say parent-report, for Construct X, say depression, often will not be perfectly correlated with the method effect for the same method for Construct Y, say extraversion, as is assumed in single-indicator MTMM models. Multiple-indicator MTMM designs have been recommended in the literature to address this issue (Eid et al., 2003; 2008; Geiser et al., 2012; Marsh, 1993; Marsh & Hocevar, 1988). Multiple-indicator MTMM designs allow for *trait-specific* method effects, thus relaxing the assumption of perfectly correlated method effects across traits (for a recent application, see Grigorenko, Geiser, Slobodskaya, & Francis, 2010).

A Basic True Score Model for MTMM Data

The original Campbell and Fiske (1959) MTMM approach uses observed correlations that are not corrected for the unreliability of the measured variables. One of the main arguments for using CFA models to analyze MTMM data is that these models allow separating measurement error from true individual differences for each TMU, thus allowing for a more accurate estimation of convergent and discriminant validity. Classical test theory (CTT; Lord & Novick, 1968) is a psychometric measurement theory that provides clear mathematical definitions of true score and error latent variables and allows formulating testable measurement models that can be specified within the CFA framework (Novick, 1966; Steyer, 1989; Zimmermann, 1975).

In agreement with Castro-Schilo et al., our starting point is the completely unrestricted assumption that each measured variable in an MTMM matrix (1) represents a TMU and (2) should be seen as fallible, that is, as potentially containing measurement error variance. In line with CTT, we can decompose each measured variable Y_{jk} in a single-indicator MTMM design (j = trait or construct, k = method) into a true score and an error variable:

$$\begin{aligned} Y_{jk} &= \tau_{jk} + \varepsilon_{jk}, \text{ where} \\ \tau_{jk} &:= E(Y_{jk}|U), \\ \varepsilon_{jk} &:= Y_{jk} - E(Y_{jk}|U), \end{aligned} \quad (1)$$

and U represents the observational unit variable, the values of which are the observational units (in MTMM studies typically persons; Steyer, 1989). That is, in defining the true score variable in CTT, we condition on the persons. In other words, in CTT, a person's true score is defined as the mean of that person's hypothetical intra-individual score distribution.

Note that the true score variable τ_{jk} is method-specific. That is, it represents the error-free scores of measured variable Y_{jk} measuring trait j (say, depression) with method k (say, parent report). The error variable ε_{jk} has an expected value of zero and is uncorrelated with τ_{jk} by definition of ε_{jk} as a regression residual in CTT (Steyer, 1989).

It follows that in this completely unrestricted decomposition, each TMU has its own true score variable. If we could measure without random measurement error (or if we knew each measure's true score and error variance components), then we could assess the "true" (in the sense of "free of measurement error") correlations between TMUs in line with Campbell and Fiske's (1959) guidelines. The basic model of CTT is a widely accepted model for conceptualizing true scores and error scores in psychology and the social sciences.

Unfortunately, test scores with perfect reliabilities or known true score and error variance are rarely found in practice. Nonetheless, we can identify the latent true score variables by using multiple (at least two) indicators per TMU as shown below. If we have multiple indicators for each TMU and make the assumption that all indicators pertaining to the same TMU share a *common* true score variable (assumption of tau-congenerity in CTT, e.g., Steyer, 1989), we can set up the following basic measurement model for multiple-indicator MTMM data (assuming mean-centered observed and latent variables for simplicity):

$$Y_{ijk} = \gamma_{ijk} \tau_{jk} + \varepsilon_{ijk}, \quad (2)$$

where γ_{ijk} indicates a constant factor loading parameter and $\gamma_{ijk} > 0$. Note that we have added an additional index i to indicate the measured variable, given that we now assume multiple measured variables per TMU. Figure 1A shows the basic MTMM measurement model for a 3(measured variables) $\times$ 3(traits) $\times$ 3(methods) MTMM design, assuming uncorrelated error variables.

Following Marsh and Hocevar (1988), we argue that the basic MTMM measurement model shown in Figure 1A represents a plausible representation of the structure of MTMM data that is in line with both the theoretical concepts of true score and error variables in CTT and Campbell and Fiske's (1959) concept of TMUs in MTMM data structures. We further argue that this model is useful for practical applications, because it implies a simple CFA model that allows MTMM researchers to estimate the latent correlations among TMUs. These correlations can be summarized in a latent MTMM matrix, which (1) can be interpreted fully in line with Campbell and Fiske's (1959) guidelines for determining convergent and discriminant validity and (2) has the additional benefit of being corrected for measurement error, thus providing a more accurate estimation and easier comparison of MTMM correlations across TMUs.

Note that the latent variable (or structural) part of the model in Figure 1A is saturated (all latent variances and covariances are freely estimated). This is in line with Campbell and Fiske's (1959) MTMM correlation matrix, which also represents a saturated structure. For F mean-centered (but not standardized) TMU factors, the model in Figure 1A estimates $[F(F + 1)]/2$ latent variances and covariances (in our example $9 \times 10/2 = 45$ free unstandardized structural parameters). We argue that it is not necessary or useful for any more sophisticated CFA-MTMM model to estimate more than this number of structural model parameters, because any additional parameter would represent an overparameterization of Campbell and Fiske's MTMM correlation structure. Next, we show that the CT-C(M – 1) model can be formulated as a straightforward reparameterization of the basic MTMM measurement model that implies exactly the same data structure as the former model, but focuses on specific contrasts between TMUs. Later on, we show that two different versions of the CT-CM model for the same design both estimate more than the number of required structural parameters derived above.

The CT-C(M – 1) Model

As Castro-Schilo et al. note in their article, in contrast to the CT-CM model, the CT-C(M – 1) model does not specify general trait factors or method factors for each method. Instead, the CT-C(M – 1) model uses as many trait factors as there are different constructs in the study, but only $M - 1$ method factors, where M indicates the total number of methods per trait. The trait and method factors in the CT-C(M – 1) model are defined—and have to be interpreted—relative to a comparison standard, the so-called *reference method*. (For a detailed discussion and guidelines as to the choice of the reference method in practical applications, see Geiser, Eid, & Nussbeck, 2008.)

Formally, the trait factors in the CT-C($M - 1$) model are defined as the true score variables pertaining to the reference method (for the formal definitions, see Appendix A). The method factors in the CT-C($M - 1$) model are defined as latent (true score) regression residuals in latent regressions on the reference method true score variables (Eid, 2000; Eid, Lischetzke, Nussbeck, and Trierweiler, 2003; Eid, Nussbeck, Geiser, Cole, Gollwitzer, & Lischetzke, 2008; Geiser et al., 2008; Geiser, Eid, West, Lischetzke, & Nussbeck, 2012).

A version of the CT-C($M - 1$) model can be derived from the basic MTMM measurement model in three steps (Geiser et al., 2012). Note that the version of the CT-C($M - 1$) model presented here does not change in fit when we change the reference method¹. In the first step, one method is selected as reference for each trait. In the second step, the true score variables pertaining to the non-reference methods are regressed on the true score variables pertaining to the reference method for each TMU (see Figure 1B; in this case, the first method [$k = 1$] was selected as reference method). The residuals of these latent regressions are the method factors. The method factors thus represent the portions of the non-reference true score variables from which the variance shared with the reference true score variables has been partialled out.

In the third step, we apply a Schmid-Leiman transformation (Schmid & Leiman, 1957) to rearrange the latent variables in such a way that the model is more in line with the way the CT-C($M - 1$) model is usually depicted in the literature (see Figure 1C). Note that all three models (1A through 1C) are equivalent and imply exactly the same data structure. Of note, the CT-C($M - 1$) model shown in Figures 1B and 1C perfectly reproduces the structure implied by the basic MTMM measurement model in Figure 1A with the same number of parameters (see Appendix B).

In the CT-C($M - 1$) model, all trait factors can be correlated and all method factors can be correlated. Correlations between trait and method factors are also allowed, except for trait and method factors pertaining to the same TMU (Eid et al., 2003). Regardless of the choice of the reference method, the resulting CT-C($M - 1$) reparameterizations of the basic MTMM measurement model will show an identical fit for a given data set and will also show an identical fit to the basic MTMM measurement model (Geiser et al., 2012).

We argue here that if we accept the fact that the basic MTMM measurement model is meaningful for MTMM data, then the CT-C($M - 1$) approach is not less, but *more* plausible for representing this data structure than the CT-CM model. The reason is that the CT-CM model overfits the data structure implied by the basic MTMM measurement model—potentially causing method factors to collapse, a problem often encountered in empirical applications of this model. This can be seen from Figures 1D and 1E, which show two different ways of specifying a CT-CM model for multiple-indicator MTMM designs. Figure 1D shows a hierarchical (second-order) CT-CM structure, whereas Figure 1E represents a CT-CM structure similar to a bifactor approach. Note that the structures in Figure 1D and 1E

¹Other versions of the CT-C($M - 1$) model have been presented in the literature that are slightly more highly parameterized and that imply a slightly different covariance structure for different choices of the reference method (e.g., Eid, 2000; Eid et al., 2003).

are not equivalent and that they are not equivalent to either the basic MTMM measurement model or the CT-C(M – 1) model.

The CT-CM model version in Figure 1D estimates a total of 102 free parameters for the present $3 \times 3 \times 3$ MTMM design, whereas the CT-CM model version in Figure 1E estimates a total of 120 free parameters. The basic MTMM measurement model as well as any version of the CT-C(M – 1) model only estimate 90 parameters for the same structure (we provide a list of all parameters for each model in Appendix B). We note that it is true in general that the CT-CM model uses more parameters to explain an MTMM data structure than the CT-C(M – 1) model.

Regardless of the choice of the reference method, the CT-C(M – 1) model estimates only 45 structural parameters for the present design—the same number of structural parameters that is estimated in the basic MTMM measurement model. In contrast, the CT-CM model version in Figure 1D estimates 57 and the version in Figure 1E estimates 51 structural parameters. This shows that either version of the CT-CM model overfits the latent TMU structure, estimating more parameters in the structural model than what is needed to fit a (saturated) MTMM covariance structure. This is obvious from Figures 1D and 1E, in which in addition to nine correlated method factors, three correlated “traits” are fit. We may ask the question what these “trait” factors measure, given that the latent MTMM covariance structure is already saturated by the nine correlated “method” factors.

Our main point here can be summarized as follows: We believe that few researchers would argue against the basic MTMM measurement model presented in Figure 1A as being in line with Campbell and Fiske’s (1959) MTMM structure. This model is just a true-score version of the classical Campbell and Fiske MTMM structure that allows researchers to correct each TMU score for measurement error (unreliability) and to evaluate Campbell and Fiske’s (1959) criteria for convergent and discriminant validity based on latent correlations (i.e., correlations corrected for measurement error). The CT-C(M – 1) model represents just a reformulation of the basic MTMM measurement model. So why should the CT-C(M – 1) model be seen as a less plausible model for representing MTMM data than the basic MTMM measurement model?

The CT-C(M – 1) model represents the same saturated structure with different parameters. That is, it allows analyzing contrasts between different methods. The CT-C(M – 1) model defines trait and method factors with regard to a reference method. As a consequence, the CT-C(M – 1) model will yield different associations with external variables, depending on the choice of the reference method (for a detailed discussion of this issue, see Geiser et al., 2008). This, however, is not surprising or confusing once one understands the definitions of the latent variables.

To understand this issue better, consider regression analysis with code variables commonly used to represent nominal independent variables in regression. In regression analysis with code variables, different coding schemes (e.g., dummy, effect, and contrast coding), if properly applied, all represent the underlying data structure equally well and yield the same overall model fit in terms of R^2 (Cohen, Cohen, West, & Aiken, 2003). Although dummy,

effect, and contrast coding schemes all result in the same overall R^2 , each of these methods yields different parameter estimates (i.e., regression coefficients) for a given set of data, because different coding schemes provide different effect decompositions. Any parameterization represents an equally valid representation of the data at hand. It depends on the researcher's goals which parameterization is most meaningful and interpretable.

In order to understand and correctly interpret the regression coefficients obtained from, for example, dummy coding, the researcher must know (1) which group was chosen as reference and (2) how the regression coefficients for dummy codes are properly interpreted. Furthermore, the researcher cannot expect to obtain identical results if he or she used (a) a different reference group or (b) a different coding scheme. This is because different reference groups and/or coding schemes necessarily yield different regression coefficients. Yet, we believe that nobody would conclude that, for example, effect coding yields more biased parameter estimates than dummy or contrast coding or that one of the coding schemes is less plausible as a "data-generating" mechanism.

A similar argument holds for the CT-C(M – 1) model: The model leads to different sets of parameters (1) compared to the basic MTMM measurement model and (2) for different choices of the reference method. However, each parameterization implies and reproduces the exact same basic MTMM measurement structure. Different model versions just explain the structure differently, as they focus on a specific user-defined set of contrasts. In the case of the basic MTMM measurement model, common true score variables as defined in CTT are estimated for each TMU and a (saturated) latent MTMM correlation structure among them can be analyzed in line with Campbell and Fiske's (1959) recommendations. In the case of a CT-C(M – 1) parameterization, a specific user-defined set of contrasts between different TMUs is analyzed. The structural (true score or latent variable) model is still saturated, that is, it perfectly reproduces the latent MTMM correlation structure.

Other Approaches to Decomposing a Latent MTMM Structure

Similar to various coding schemes in regression, multiple different ways of decomposing the latent MTMM structure have been presented in the literature. For example, instead of using a regression approach to contrasting different TMU true score variables as in the CT-C(M – 1) model, a latent difference parameterization can also be used, in which the method factors are defined in terms of latent difference scores relative to a reference method (see Figure 2; Geiser et al., 2012; Pohl, Steyer, & Kraus, 2008; formal definitions are provided in Appendix A). Researchers interested in studying method effects as deviations from "common latent traits" rather than from "reference-method based traits" can define latent trait factors as the grand mean of the true score variables pertaining to the same trait j , leading to an effect coding scheme as discussed in Pohl and Steyer (2010; see Figure 3; formal definitions are provided in Appendix A). All these parameterizations have in common that they (1) are equivalent to the basic MTMM measurement model, (2) provide a complete decomposition of the basic MTMM true score structure, and (3) use constructive definitions of trait and method latent variables.

By *constructive* we mean that these parameterizations explicitly define trait and method latent variables in terms of true score variables or mathematical functions of true score

variables. The trait and method factors are “constructed” out of a set of clearly defined latent true score variables by introducing explicit mathematical definitions for traits and methods. In the CT-C(M – 1), latent difference, and latent means approaches to MTMM modeling, we therefore know exactly what the “trait” and “method” factors are and how they are properly interpreted. In contrast, the trait and method factors in the CT-CM model cannot be derived through explicit mathematical definitions on the basis of the TMU true score variables in Figure 1A. Therefore, it is less clear what these factors really are and how they should be interpreted.

Bias in the CT-C(M – 1) Model?

Castro-Schilo et al. speak of “bias” in the context of their simulation when they compare the estimated parameters in the CT-CU and CT-C(M – 1) approaches to the population parameters used to generate data based on the CT-CM model. In this section, we elaborate more on why we think that the term “bias” may be misleading for results obtained from CT-C(M – 1) analyses of data generated by the CT-CM model and how the results from this model are properly interpreted. We focus on the CTC(M – 1) model, because we feel that many researchers in the field are still confused about how to properly interpret results from this model.

As a consequence of the different definitions of trait and method factors in the CT-C(M – 1) model, it is logical that the CT-C(M – 1) model will yield different results when fit to data generated from the CT-CM model. Geiser et al. (2008) discussed the reasons for why the CT-C(M – 1) approach generally results in different estimates of trait and method factor associations with external variables than other CFA-MTMM approaches in detail.

In the CT-CM model, trait factors are conceptualized as “general” (i.e., method un-specific) factors. In addition, method factors are included for each method in the CT-CM model. In contrast, in the CT-C(M – 1) model, trait factors are by definition *method specific*, because they are defined as the common true score variables of the reference method (e.g., true self-reported depression). Method factors are defined and have to be interpreted relative to the reference true score variable in the CTC(M – 1) model. As a consequence, interpretations of associations between trait and method factors with external variables differ by definition between the CT-CM and CT-C(M – 1) models. We argue that one should not interpret these differences as bias. Rather, each model reflects the associations in a different way. In fact, we argue that the CT-C(M – 1) model provides results that are easier to interpret, because we know exactly how the trait and method factors are defined in this model.

If we relate, for example the reference factor in the CT-C(M – 1) model (say self-reported depression) to an external variable (say, school grades), then we know that this association reflects the association between the true (error-free) depression self-rating scores and school performance. We furthermore know that the association of the method factor for, say mother reports of depression, with school grades represents the semi-partial correlation between the unique aspects of the true mother ratings (above and beyond what true mother ratings share with true self-ratings) with school grades. The question of what the associations between the “general traits” and method factors in the CT-CM model and external variables mean is less

clear, because the trait and method factors in the CT-CM model cannot be derived as direct functions of well-defined true score variables.

As we explained above, different choices of the reference method in the CT-C(M – 1) model all imply the same overall covariance structure, but yield different results in terms of parameter estimates and associations with external variables, because the meaning of the reference factor and method factors changes, depending on the choice of the reference method. For example, in one application, we may find it most useful to contrast self-ratings of depression with mother, father, and teacher ratings of the same construct. In another application, the focus may be more on contrasting teacher ratings against self-, mother, and father reports. Or, there may be evidence for a particular method as being a more valid (“gold standard”) causing a researcher to select this method as reference (e.g., objective measures of intelligence vs. self- and other ratings).

The question of which method should be chosen as reference is similar to the issue of choosing a coding scheme in regression analysis. It depends on the goals of the researcher and the question of which set of contrasts is most meaningful in the study at hand. It is therefore important that researchers choose the reference method in the CT-C(M – 1) or other MTMM models (e.g., the latent difference model) based on theoretical or substantive grounds in line with their specific research questions and hypotheses (Geiser et al., 2008). None of the contrasts, however, is more “biased” than another. Different contrasts simply provide researchers with views of the data from different angles.

Discussion

Analyzing MTMM data is a complex endeavor. We fully agree with Castro-Schilo et al. (2013) that the question of which MTMM model is the most suitable model “cannot be answered easily or definitively” (p. 205). However, we think that a strong case can be made for CFA-MTMM models that provide (1) complete decompositions of the latent MTMM covariance structure and (2) clear mathematical definitions of trait and method factors based on the well-defined true score variables derived from CTT. In this comment, we laid out our arguments against the suggestion that the CT-CM model is the theoretically most attractive model for MTMM studies and that the CT-C(M – 1) model yields biased results relative to the “true” MTMM structure. We furthermore provided arguments against Castro-Schilo et al.’s claim according to which “MTMM structural models that do not specify all MVs explicitly as decomposable trait-method units partially ignore the true structure of the data.” (p. 204). We showed that the CT-C(M – 1) and other CFA-MTMM models fully reproduce the basic MTMM measurement structure of true score variables. In contrast, the CT-CM model overfits this structure with factors that are partly superfluous, explaining the high rate of empirical and interpretation problems seen for this model. We believe that if one needs to choose a general CFAMTMM model for a simulation study, the model in Figure 1A would be the natural choice, because this model appropriately represents Campbell and Fiske’s (1959; Fiske & Campbell, 1992) ideas (see discussion below).

We argue that it is usually impossible for researchers in real data situations to know which model or mechanism generated the data, other than that we may know which *types* of

methods were used in the study (see Eid, Nussbeck, Geiser, Cole, Gollwitzer, & Lischetzke, 2008, or Nussbeck, Eid, Geiser, Courvosier & Lischetzke, 2009, for details on the distinction between random [interchangeable] and fixed [structurally different] methods in MTMM designs and implications of this distinction for the choice of an appropriate CFA-MTMM model). In our view, what the field needs are psychometric theories about how “trait”, “method”, and “error” latent variables can be constructively and meaningfully defined conceptually and mathematically, that is, defined in ways that we know what exactly these latent variables are and how they are properly interpreted.

We suggest that probably the best that MTMM researchers can do is start out with a basic measurement theory such as CTT to separate random measurement error from true TMU variance in MTMM studies and then think about which effect decomposition of the basic MTMM measurement structure provides the most meaningful answers to the questions asked in the study. Using CFAMTMM approaches that reparameterize the basic MTMM measurement model (Figure 1A) in line with the researcher’s goals will have the benefit of dealing with (1) well-defined latent trait and method variables that are functions of well-defined true score variables from CTT and have clear meanings and interpretations and (2) models that are likely to provide fewer convergence and estimation problems than the CT-CM model, because these models represent just reparameterizations of the common factor model shown in Figure 1A.

The CT-C(M – 1) model is one approach that meets this goal. This model is based on the theoretically well-founded and widely accepted true score decomposition in CTT. As all latent trait and method factors in this approach are functions of well-defined true score-variables, these variables as well as their associations with each other and with external variables have clear meanings and interpretations. Other CFA-MTMM approaches that also define trait and method factors constructively based on true score variables and that provide a full representation of the basic MTMM measurement model are Pohl et al.’s (2008) latent difference method factor approach as well as Pohl and Steyer’s (2010) latent means model.

***M* Versus *M* – 1 Correlated Method Factor Approaches**

All three approaches [CT-C(M – 1), latent difference, and latent means approach] have in common that they provide different ways of defining effects based on the basic MTMM measurement model. This idea parallels the definition of effects in analysis of variance (ANOVA) and regression analysis with code variables. For example, we distinguish between dummy coding [which uses a reference approach similar to the CT-C(M – 1) and latent difference MTMM models] and effect coding (which uses a “deviation from the grand mean” approach similar to what is done in the latent means MTMM model discussed by Pohl & Steyer, 2010).

Interestingly, a common feature of the CT-C(M – 1), latent difference, and latent means approaches to MTMM analysis is that they all require *T* trait, but only *M* – 1 correlated method factors to fully represent the basic MTMM measurement structure. To us, this is another sign that models with *M* correlated method factors such as the CT-CM model generally overparameterize MTMM data. In our view, extracting as many correlated trait factors as there are constructs and at the same time as many correlated method factors as

there are methods in the study means not only asking too much from the data—doing so is simply unnecessary. This issue parallels the fact that in regression analysis with code variables $G - 1$ codes are enough to fully represent the data structure (where G indicates the total number of groups of the independent variable). If we used an additional code variable (i.e., a total of G codes), the model would be overparameterized. In the last section, we argue that our proposed approach of using reparameterizations of the basic MTMM measurement model like the CT-C($M - 1$) model is not only more parsimonious than the CT-CM model, but also fully in line with Fiske and Campbell's (1992) late view of their MTMM approach.

Fiske and Campbell's (1992) Late View of the MTMM Approach

In a review of their seminal work on the MTMM matrix and its impact on psychology over the years, Fiske and Campbell (1992) stated that “perhaps methods and traits or contents are so thoroughly intertwined that their interaction cannot be adequately analyzed” (p. 393) and that “Method and trait or content are highly interactive and interdependent. We may have to settle for the practice of studying ‘trait method units’ (as we called them in the original [Campbell & Fiske, 1959] article) if we can do so without entering the blind alley of pure operationalism: having a distinct operation for each such combination” (p. 394).

We think that the CT-C($M - 1$), latent difference, and latent means approaches are in full accordance with this idea: these models provide tools to study TMUs in ways that depend on the researcher's goals. For example, in some cases, it may be most interesting or useful to study contrasts of specific methods against a reference method. This could be the case, for instance, when a researcher has a gold standard measure (e.g., an objective measure of intelligence) and wants to contrast this measure against other, less established measures of the same construct (e.g., teacher, parent, and self-ratings of intelligence) to find out whether other methods that may, for example, be cheaper to employ, provide similar results. In other cases, researchers may be more interested in defining a common trait for each construct and analyzing method effects as deviations from the average across methods (Pohl & Steyer, 2010).

Our point is that MTMM modeling should involve explicit definitions of trait and method factors on the basis of the researchers' goals—in much the same way as we specify user-defined contrasts in ANOVA or regression. If the researcher takes on this responsibility, then the typically unresolvable question of which mechanism or model generated the data will become less relevant, because the focus will shift towards better understanding the way different methods or raters work in the assessment of psychological constructs.

Conclusion

Studying different methods for analyzing MTMM data is important for clarifying theoretical and empirical differences between different approaches, and we appreciate Castro-Schilo et al.'s (2013) efforts in this regard. In this comment, we have clarified our perspective on some of the assumptions and interpretations in Castro-Schilo et al.'s paper. We hope that researchers will find our alternative perspective on MTMM analyses, which is based on

meaningful decompositions of clearly defined latent true score variables, to be useful as they think about how to best approach their own CFAMTMM analyses.

Acknowledgments

This research was funded in part by grant 1 R01 DA034770-01 from the National Institutes on Drug Abuse (NIH-NIDA) awarded to Christian Geiser and Ginger Lockhart.

The authors would like to thank Ginger Lockhart, Roxanne Felser, and Natalee Greyhart for helpful comments on the draft.

Appendix A: Derivation of the CT-C(M – 1), Latent Difference, and Latent Means Models for MTMM Data

In this appendix, we show how the trait and method factors in the CT-C(M – 1) (Geiser et al., 2008; 2012), latent difference (Pohl, Steyer, & Kraus, 2008; Geiser et al., 2012), and latent means models (Pohl & Steyer, 2010) can be constructively defined based on the common true score variables in the basic MTMM measurement model (Figure 1A). The measurement equation of the basic MTMM measurement model (Equation 2 in the main text) is repeated here as Equation A1 for convenience:

$$Y_{ijk} = \lambda_{ijk} \tau_{jk} + \varepsilon_{ijk}, \quad (\text{A1})$$

where γ_{ijk} indicates a constant factor loading parameter and $\gamma_{ijk} > 0$. For simplicity, we assume that all variables are centered. Note that all models presented here can also be formulated for uncentered variables (e.g., Geiser et al., 2012; Pohl & Steyer, 2010).

CT-C(M – 1) Model

In the CT-C(M – 1) model, one method per trait serves as reference. Without loss of generality, we select the first method ($k = 1$) as reference. The true score variables τ_{j1} pertaining to the reference method serve as (method-specific) trait factors in the CT-C(M – 1) model. The remaining methods are contrasted against the reference method in a latent regression analysis (see Figure 1B):

$$\tau_{jk} = \beta_{jk} \tau_{j1} + M_{jk}, \quad (\text{A2})$$

where β_{jk} indicates a constant latent regression slope coefficient and the trait-specific method factors M_{jk} are defined as residuals of this regression

$$M_{jk} := \tau_{jk} - E(\tau_{jk} | \tau_{j1}). \quad (\text{A3})$$

Inserting Equation A2 into Equation A1 yields the model for the measured variables shown in Figure 1C:

$$\begin{aligned} Y_{ijk} &= \lambda_{ijk} \tau_{jk} + \varepsilon_{ijk} \\ &= \lambda_{ijk} (\beta_{jk} \tau_{j1} + M_{jk}) + \varepsilon_{ijk} \quad (\text{A4}) \\ &= \lambda_{ijk} \beta_{jk} \tau_{j1} + \lambda_{ijk} M_{jk} + \varepsilon_{ijk}. \end{aligned}$$

In the CT-C($M - 1$) model, method and trait factors pertaining to the same trait j are not allowed to correlate, because the method factors M_{jk} are defined as residuals with respect to τ_{j1} . All other latent correlations are allowed in the model.

Latent Difference Model

The latent difference approach (Figure 2) also contrasts $M - 1$ methods against a reference method for each trait. Without loss of generality, we again select the first method ($k = 1$) as reference. The true score variables τ_{j1} pertaining to the reference method serve as (method-specific) trait factors in the latent difference model. The remaining methods are contrasted against the reference method by including latent difference variables (Figure 2A):

$$\begin{aligned}\tau_{jk} &= \tau_{j1} + (\tau_{jk} - \tau_{j1}) \\ &= \tau_{j1} + M_{jk},\end{aligned}\quad (\text{A5})$$

where the trait-specific method factors M_{jk} are defined as the deviation (latent difference) from the reference method:

$$M_{jk} := (\tau_{jk} - \tau_{j1}). \quad (\text{A6})$$

Inserting Equation A5 into Equation A1 yields the model for the measured variables shown in Figure 2B:

$$\begin{aligned}Y_{ijk} &= \lambda_{ijk} \tau_{jk} + \varepsilon_{ijk} \\ &= \lambda_{ijk} [\tau_{j1} + (\tau_{jk} - \tau_{j1})] + \varepsilon_{ijk} \\ &= \lambda_{ijk} \tau_{j1} + \lambda_{ijk} (\tau_{jk} - \tau_{j1}) + \varepsilon_{ijk} \\ &= \lambda_{ijk} \tau_{j1} + \lambda_{ijk} M_{jk} + \varepsilon_{ijk}.\end{aligned}\quad (\text{A7})$$

In the latent difference model, all method and trait factors are allowed to correlate. Note that the latent difference model requires the scores of different methods to be in the same units of measurement for a meaningful interpretation of the results (Geiser et al., 2012). This assumption can be tested by checking the fit of a model that holds the factor loadings (and, if included, intercepts) constant across methods for the same indicator.

Latent Means Model

In the latent means approach (Figure 3) common trait factors T_j are constructively defined as the average across the true score variables pertaining to the same trait. For three methods per trait, we obtain:

$$T_j := (\tau_{j1} + \tau_{j2} + \tau_{j3}) / 3. \quad (\text{A8})$$

Trait-specific method factors M_{jk} are defined as deviations (difference scores) from the average factor defined in Equation A8:

$$M_{jk} := \tau_{jk} - T_j. \quad (\text{A9})$$

As a consequence of their definition as deviations from the mean, the method factors sum to zero within each trait:

$$M_{j1} + M_{j2} + M_{j3} = 0. \quad (\text{A10})$$

Because of the implicit sum-to-zero constraint, one method factor per trait is redundant, as can be seen from the following relationships between the method factors that follow directly from Equation A10:

$$\begin{aligned} M_{j1} &= -M_{j2} - M_{j3} \\ M_{j2} &= -M_{j1} - M_{j3} \\ M_{j3} &= -M_{j1} - M_{j2}. \end{aligned} \quad (\text{A11})$$

We therefore need to estimate only $M - 1$ method factors per trait. If, without loss of generality, we decide to drop the first method factor (M_{j1}), we can write the true score variables as illustrated in Figure 3A:

$$\tau_{j1} = T_j - M_{j2} - M_{j3} \quad (\text{A12})$$

$$\tau_{j2} = T_j + M_{j2} \quad (\text{A13})$$

$$\tau_{j3} = T_j + M_{j3}. \quad (\text{A14})$$

Inserting Equations A12-A14 into Equation A1 yields the equations for the measured variables illustrated in Figure 3B:

$$\begin{aligned} Y_{ij1} &= \lambda_{ij1} \tau_{j1} + \varepsilon_{ij1} \\ &= \lambda_{ij1} [T_j - M_{j2} - M_{j3}] + \varepsilon_{ij1} \\ &= \lambda_{ij1} T_j - \lambda_{ij1} M_{j2} - \lambda_{ij1} M_{j3} + \varepsilon_{ij1}. \end{aligned} \quad (\text{A15})$$

$$\begin{aligned} Y_{ij2} &= \lambda_{ij2} \tau_{j2} + \varepsilon_{ij2} \\ &= \lambda_{ij2} [T_j + M_{j2}] + \varepsilon_{ij2} \\ &= \lambda_{ij2} T_j + \lambda_{ij2} M_{j2} + \varepsilon_{ij2}. \end{aligned} \quad (\text{A16})$$

$$\begin{aligned} Y_{ij3} &= \lambda_{ij3} \tau_{j3} + \varepsilon_{ij3} \\ &= \lambda_{ij3} [T_j + M_{j3}] + \varepsilon_{ij3} \\ &= \lambda_{ij3} T_j + \lambda_{ij3} M_{j3} + \varepsilon_{ij3}. \end{aligned} \quad (\text{A17})$$

In the latent means model, all method and trait factors are allowed to correlate. Note that the latent means model requires the scores of different methods to be in the same units of measurement for a meaningful interpretation of the results (Pohl & Steyer, 2010). This assumption can be tested by checking the fit of a model that holds the factor loadings (and, if included, intercepts) constant across methods for the same indicator.

Appendix B: Numbers of Parameters in the Basic MTMM Measurement, CT-C(M – 1), CT-CM, Latent Difference, and Latent Means Models for Multiple-Indicator MTMM Data

Table B1

Overview of Free Parameters Estimated in the CFA-MTMM Models Shown in Figures 1-3 for a 3(Measured Variables) $\times$ 3(Traits) $\times$ 3(Methods) MTMM Design

Parameter Type	Basic MTMM Measurement Model (Figure 1A)	CT-C(M – 1) (Figure 1B)	CT-CM (Figure 1D)	CT-CM (Figure 1E)	Latent Difference (Figure 2A)	Latent Means (Figure 3A)
Measurement model						
True score loadings λ_{ijk}	18	18	18	—	18	18
Trait factor loadings λ_{Tijk}	—	—	—	24	—	—
Method factor loadings λ_{Mijk}	—	—	—	18	—	—
Error variances $\text{Var}(\varepsilon_{ijk})$	27	27	27	27	27	27
Structural model						
True score variances $\text{Var}(\tau_{jk})$	9	(3) ^a	—	—	(3) ^a	—
True score covariances	36	—	—	—	—	—
Latent regression coefficients β_{jk}	—	6	—	—	—	—
Trait factor loadings λ_{Tjk}	—	—	6	—	—	—
Trait factor variances $\text{Var}(\tau_{j1})$ or $\text{Var}(T_j)$	—	3	3	3	3	3
Method factor variances $\text{Var}(M_{jk})$	—	6	9	9	6	6
Trait-Trait covariances	—	3	3	3	3	3
Trait-Method covariances	—	12	—	—	18	18
Method-Method covariances	—	15	36	36	15	15
Sum (measurement model)	45	45	45	69	45	45
Sum (structural model)	45	45	57	51	45	45
Sum (total)	90	90	102	120	90	90

Note. We assume mean-centered variables. We furthermore assume that the metric of each factor is assigned by fixing one loading per factor to 1.0. All remaining loadings and all latent variances are assumed to be freely estimated.

^aIn the CT-C(M – 1) and latent difference models, the trait factors are defined as the true score variables τ_{j1} pertaining to the reference method ($k = 1$). Hence, three true score variances $\text{Var}(\tau_{j1})$ are estimated in these models, respectively. Here, we count these variances under “trait variances” to facilitate the comparison between models.

References

- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Castro-Schilo L, Widaman KF, Grimm KJ. Neglect the structure of multitrait-multimethod data at your peril: Implications for associations with external variables. *Structural Equation Modeling*. 2013; 20:181–207.
- Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. Erlbaum; Mahwah, NJ: 2003.
- Eid M. A multitrait-multimethod model with minimal assumptions. *Psychometrika*. 2000; 65:241–261.
- Eid M, Lischetzke T, Nussbeck FW, Trierweiler LI. Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple indicator CT-C(M-1) model. *Psychological Methods*. 2003; 8:38–60. [PubMed: 12741672]
- Eid M, Nussbeck FW, Geiser C, Cole DA, Gollwitzer M, Lischetzke T. Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*. 2008; 13:230–253. [PubMed: 18778153]
- Fiske DW, Campbell DT. Citations do not solve problems. *Psychological Bulletin*. 1992; 112:393–395.
- Geiser C, Eid M, Nussbeck FW. On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares & Coffman (2006). *Psychological Methods*. 2008; 13:49–57. [PubMed: 18331153]
- Geiser C, Eid M, West SG, Lischetzke T, Nussbeck FW. A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling*. 2012; 19:409–436.
- Grayson D, Marsh HW. Identification with deficient rank loading matrices in confirmatory factor analysis: Multitrait-multimethod models. *Psychometrika*. 1994; 59:121–134.
- Grigorenko EL, Geiser C, Slobodskaya HR, Francis DJ. Cross-informant symptoms from CBCL, TRF, and YSR: Trait and method variance in a normative sample of Russian youths. *Psychological Assessment*. 2010; 22:893–911. [PubMed: 21133549]
- Kenny DA, Kashy DA. The analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*. 1992; 112:165–172.
- Lord, FM.; Novick, MR. *Statistical theories of mental test scores*. Addison-Wesley; Reading, MA: 1968.
- Marsh HW. Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*. 1989; 13:335–361.
- Marsh HW. Multitrait-multimethod analyses: Inferring each trait/method combination with multiple indicators. *Applied Measurement in Education*. 1993; 6:49–81.
- Marsh HW, Bailey M. Confirmatory factor analysis of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*. 1991; 15:47–70.
- Marsh HW, Hocevar D. A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*. 1988; 73:107–117.
- Novick MR. The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*. 1966; 3:1–18.
- Nussbeck FW, Eid M, Geiser C, Courvoisier DS, Lischetzke T. A CTC(M-1) model for different types of raters. *Methodology*. 2009; 5:88–98.
- Pohl S, Steyer R. Modeling common traits and method effects in multitrait-multimethod analysis. *Multivariate Behavioral Research*. 2010; 45:1–28.
- Pohl S, Steyer R, Kraus K. Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society*. 2008; 171:41–63. Series A
- Rindskopf D. Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research*. 1984; 13:109–119.

- Schmid J, Leiman J. The development of hierarchical factor solutions. *Psychometrika*. 1957; 22:53–61.
- Steyer R. Models of classical psychometric test theory as stochastic measurement models: Representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika*. 1989; 3:25–60.
- Steyer R. Das MTMM-Modell ist nicht identifiziert [The MTMM model is not identified]. *Newsletter der Fachgruppe Methoden*. 1995; 3:5.
- Zimmermann DW. Probability spaces, Hilbert spaces, and the axioms of test theory. *Psychometrika*. 1975; 40:395–412.

Figure 1.

Transforming a common factor model for a 3(measured variables) × 3(traits) × 3(methods) MTMM design into a CT-C(M – 1) model. A: Basic MTMM measurement model with correlated true score variables representing the TMUs. B: Reparameterizing Model A as a latent regression [CT-C(M – 1)] model using Method 1 ($k = 1$) as reference method for each trait. C: Rearranging Model B to be in line with the common representation of the CT-C(M – 1) model in the literature. Figure 1A-1C show equivalent models. D: CT-CM as hierarchical (second-order) model for the same MTMM design. E: CT-CM model as bifactor-type model for the same MTMM design.

Figure 2. Reparameterizing the basic MTMM measurement model from Figure 1A as a latent difference model. A: Latent difference score formulation using Method 1 ($k = 1$) as reference method for each trait. B: Rearranging the model so that the measured variables load directly onto the trait and method factors. Figure 2A and B show equivalent models. Note that the latent difference model also requires only $M - 1$ (rather than M) correlated method factors per construct. All trait and method factors in this model are allowed to correlate.

Figure 3.

Reparameterizing the basic MTMM measurement model from Figure 1A as a latent means model. The trait factors are common traits that represent the average across methods for each trait. A: Decomposition of true score variables. B: Rearranging the model so that the measured variables load directly onto the trait and method factors. Figure 3A and B show equivalent models. Note that the latent means model also requires only $M - 1$ (rather than M) correlated method factors per construct. All trait and method factors in this model are allowed to correlate.