Organizational Research Methods
15(1) 134-161

DOI: 10.1177/1094428111408616
http://orm.sagepub.com

An Examination of G-Theory Methods for Modeling Multitrait–Multimethod Data: Clarifying Links to Construct Validity and Confirmatory Factor Analysis

David J. Woehr[1], Dan J. Putka[2], and Mark C. Bowler[3]

Abstract
For nearly three decades, the predominant approach to modeling the latent structure of multitrait–multimethod (MTMM) data in organizational research has involved confirmatory factor analysis (CFA). Despite the frequency with which CFA is used to model MTMM data, commonly used CFA models may produce ambiguous or even erroneous results. This article examines the potential of generalizability theory (G-theory) methods for modeling MTMM data and makes such methods more accessible to organizational researchers. Although G-theory methods have existed for more than half a century, the research literature has yet to provide a clear description and integration of latent models implied by univariate and multivariate G-theory with MTMM data, notions of construct validity, and CFA. To help fill this void, the authors first provide a jargon-free overview of the univariate and multivariate G-theory models and analytically demonstrate linkages between their parameters (variance and covariance components), elements of the MTMM matrices, indices of convergent and discriminant validity, and CFA. The authors conclude with a discussion and empirical illustration of a G-theory-based modeling process that helps clarify the use of G-theory methods for modeling MTMM data.

Keywords
generalizability theory, multitrait–multimethod, confirmatory factor analysis

[1] Department of Management, University of North Carolina at Charlotte, USA
[2] Human Resources Research Organization (HumRRO), Alexandria, VA, USA
[3] Department of Psychology, East Carolina University, Greenville, NC, USA

Corresponding Author:
David J. Woehr, Department of Management, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
Email: djwoehr@carolina.rr.com

In the 50 years since its introduction, Campbell and Fiske's (1959) multitrait–multimethod (MTMM) matrix approach has become a cornerstone for evaluating construct-related validity evidence. Indeed, a cursory search using Google Scholar indicates nearly 7,500 citations of Campbell and Fiske's original article. The MTMM model has been applied to a variety of constructs of interest to organizational researchers. These constructs include, but are certainly not limited to, (a) personality and attitudes (e.g., Baltes, Bauer, Bajdo, & Parker, 2002; Mitte & Kämpfe, 2008), (b) anxiety and burnout (e.g., Cresswell & Eklund, 2006; Halbesleben & Demerouti, 2005), (c) multisource performance ratings (e.g., Hoffman & Woehr, 2009; Woehr, Sheehan, & Bennet, 2005), and (d) assessment center ratings (e.g., Bowler & Woehr, 2006; Lance, Lambert, Gewin, Lievens, & Conway, 2004). Moreover, common to the vast majority of empirical studies evaluating construct-related validity within an MTMM framework is the use of confirmatory factor analysis (CFA). However, the typical CFA model applied to MTMM data is plagued with a host of methodological problems. Specifically, it is widely recognized that reliance on the traditional CFA model for the analysis of MTMM data typically results in improper or unstable solutions, particularly as the number of latent constructs increases (Kenny & Kashy, 1992; Marsh, 1989; Marsh & Bailey, 1991; Marsh, Byrne, & Craven, 1993; Marsh & Grayson, 1995). These problems are so prevalent that Kenny and Kashy (1992) lament the fact that 30 years after the introduction of the MTMM approach, we still do not know how to statistically evaluate convergent and discriminant validity adequately.

A recent study by Lance, Woehr, and Meade (2007) helps illustrate this issue. Specifically, Lance et al. present a Monte Carlo study examining multiple "true" models postulated to underlie samples of MTMM data. For each of three population models, Lance et al. generated 500 sample MTMM matrices. These 1,500 sample matrices were then separately analyzed via CFA, with each sample matrix evaluated using parameterizations representing each of the three population models. Thus, Lance et al. were able to directly evaluate whether a CFA would produce convergent and admissible solutions when the correct model was applied to the respective sample matrix (i.e., does the "right" model fit the data?). Moreover, they were able to evaluate what happens when an incorrect model was applied to a sample matrix (i.e., does the "wrong" model fit the data?).

Results of the analyses presented by Lance et al. (2007) indicated that, regardless of whether or not the data corresponded to the model specifications used, CFAs based on models with multiple-dimension factors converged to an admissible solution for only 57% of the data matrices. Even more striking was that for all models that reached an admissible solution, regardless of whether or not the fitted model matched the population model on which the data were based, traditional model fit statistics—for example, the root mean square error of approximation (RMSEA), comparative fit index (CFI), nonnormed fit index (NNFI), expected cross-validation index (ECVI), and so forth—indicated that the model provided a good fit to the data. Based on these results, Lance et al. concluded that commonly used CFA models of MTMM data are problematic due to their propensity to produce results that conflict with the true nature of the data.

The results of Lance et al.'s (2007) study serve to highlight the serious concerns raised with commonly used CFA models in MTMM literature. In addition, they argue against the continued exclusive reliance on the CFA-based approach for evaluating internal construct validity via MTMM data and call for the use of alternative evaluation approaches. One such alternative stems from the application of generalizability theory (G-theory) methods to assess the impact of multiple facets of measurement (e.g., traits and methods; Brennan, 2001; Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Although the random-effects analysis of variance (ANOVA) model that provides the basis of G-theory was once widely recognized as a potentially fruitful approach to the evaluation of MTMM data (e.g., King, Hunter, & Schmidt, 1980; Schmitt & Stults, 1986), over the past two decades few organizational research studies have utilized this approach. In contrast, G-theory methods have remained popular in the analyses of MTMM data in educational measurement (e.g., Cronbach, Linn, Brennan, & Haertel, 1997; Shavelson, Baxter, & Gao, 1993; Webb, Schlackman, & Sugrue, 2000), in part because of such methods' ability to deal with the complex measurement designs often confronted in educational research. The decline of ANOVA-based approaches for analyzing MTMM data in organizational research coincided with the rise of CFA-based strategies in the mid-1980s. As noted by authors at the time, CFA was viewed as a method that allowed researchers to relax some of the assumptions ANOVA-based approaches such as G-theory required and to generate parameter estimates for individual trait–method units—something ANOVA-based approaches fail to provide (Schmitt & Stults, 1986). Unfortunately, as previously discussed, the added flexibility this approach provides researchers has come at a price (e.g., Lance et al., 2007), and as such it is important to reconsider the benefits of G-theory methods.

One practical difficulty that may be preventing more widespread use of G-theory methods for analyzing MTMM data is the lack of a clear framework for implementing them with MTMM data and linking them to concepts typically associated with MTMM data (e.g., MTMM correlations, construct validity, CFA). As we note in later sections, the literature that does exist in this regard has linked G-theory to these concepts in an inconsistent and piecemeal manner. This has made it difficult for organizational researchers to appreciate the connections that exist between these concepts, the ease with which the G-theory methods can be implemented, and their potential value.

Exacerbating issues with the existing literature is the jargon that typically pervades discussions of G-theory and the perceived complexity of estimating variance components that provide the basis of G-theory (DeShon, 2002; Putka & Sackett, 2010)—both of which we address in this article. Typically, G-theory is introduced and discussed in the context of reliability estimation, as it represents an extension of classical test theory for partitioning observed score variance (Brennan, 2001; Cronbach et al., 1972). Along those lines, discussions of G-theory have often been laden with idiosyncratic terminology (e.g., *universes of generalization*) that is largely unnecessary for applications to MTMM data discussed herein. Thus, the very nature of the G-theory literature and perceptions of it arguably inhibit its use among organizational researchers, especially for modeling MTMM data. As such, there appears to be a definite need for clarifying how G-theory can be capitalized upon to model MTMM data.

The Current Work

In the following, we aim to provide organizational researchers with a parsimonious illustration of how G-theory can be leveraged to model MTMM data. First, we provide a brief overview of the univariate G-theory model that is free from the idiosyncratic jargon associated with its traditional use for reliability estimation. Second, we show how G-theory variance components can be directly related to, and derived from, MTMM correlations. This in turn allows for the explicit linking of G-theory variance components to Campbell and Fiske's (1959) traditional notions of convergent and discriminant validity. Third, we show how the univariate G-theory model can be expressed as a structural model, thereby explicitly linking G-theory to a CFA framework. Fourth, we show how the univariate G-theory model can be modified to fit a multivariate G-theory model to MTMM data and the benefits and drawbacks of doing so. Finally, we discuss and empirically illustrate a G-theory-based modeling process that helps clarify the interpretation and suggested use of G-theory methods for modeling MTMM data.

Although yet to be discussed in the organizational research literature, the extension of G-theory methodology to the multivariate case is important in that when organizational researchers model MTMM data they are typically interested in multiple distinct, yet interrelated, traits—precisely the situation that multivariate G-theory was intended to handle (Brennan, 2001; Webb & Shavelson, 1981). In contrast, traditional applications of univariate G-theory focus on estimating variance components underlying measures of a single trait or construct. Our treatment of multivariate G-theory is also important in that it alerts prospective users to its similarity to correlated trait–correlated uniqueness (CT-CU) models, which have been heavily criticized in the research methods literature but have escaped the attention of G-theorists (e.g., Lance, Noble, & Scullen, 2002).

Arguably, a unified, parsimonious treatment of how one can leverage G-theory to model MTMM data has yet to be provided in the organizational research literature. As we note below, this has resulted in confusion and misunderstandings regarding the substance and potential utility of G-theory for this purpose. As such, the present work should provide a valuable aid as G-theory methods are considered by more organizational researchers looking for a complement or alternative to traditional CFA-based approaches to modeling MTMM data.

The Univariate G-Theory Model and MTMM Data

To begin our illustration, the lower diagonal of Table 1 shows a simple MTMM correlation matrix, comprising three traits (t1, t2, t3) each measured by three methods (m1, m2, m3). Researchers attempting to model these data via univariate G-theory may conceive of the data collection design underlying these data as persons (p) crossed with traits (t) crossed with methods (m), or in G-theory shorthand p $\times$ t $\times$ m. That is, in this example, each person completes measures of three traits, and each trait is assessed via the same three measurement methods.[1] The statistical model underlying univariate G-theory is a random-effects ANOVA model (Cronbach et al., 1972; Jackson & Brashers, 1994). Based on this model, each person's observed score (X_{ptm}) on a given trait–method unit is modeled as a simple additive function (Brennan, 2001):

$$X_{\text{ptm}} = \mu + \nu_{\text{p}} + \nu_{\text{t}} + \nu_{\text{m}} + \nu_{\text{pt}} + \nu_{\text{pm}} + \nu_{\text{tm}} + \nu_{\text{ptm,r}}, \tag{1}$$

where μ is the grand mean score across all persons, traits, and methods; ν_{p} is the person main effect and conceptually reflects the expected value of a person's score (expressed as a deviation from the grand mean) across the population of traits and methods; ν_{t} is the trait main effect and conceptually reflects the expected value of a trait's effect (again, expressed as a deviation from the grand mean) across the population of persons and methods; ν_{m} is the method main effect and conceptually reflects the expected value of a method's effect (again, expressed as a deviation from the grand mean) across the population of persons and traits; ν_{pt} is the Person $\times$ Trait interaction effect and conceptually reflects differences in the ordering of persons' expected scores (averaged over methods) across traits; ν_{pm} is the Person $\times$ Method interaction effect and conceptually reflects differences in the ordering of persons' expected scores (averaged over traits) across methods; ν_{tm} is the Trait $\times$ Method interaction effect and conceptually reflects differences in the ordering of traits' expected scores (averaged over persons) across methods; and finally, $\nu_{\text{ptm,r}}$ is the remaining residual after accounting for all other effects in the model.[2]

The assumptions underlying this model reflect common random-effects ANOVA assumptions. Namely, all effects in the model are assumed to be independently and identically distributed with means of zero and variances of σ^2_{p}, σ^2_{t}, σ^2_{m}, σ^2_{pt}, σ^2_{pm}, σ^2_{tm}, and $\sigma^2_{\text{ptm,r}}$, respectively (Jackson & Brashers, 1994; Searle, Casella, & McCulloch, 1992). It is these *variance components* that are the focus of estimation efforts in G-theory. As noted, each term in the model (i.e., each main effect term, each interaction term, and the residual term) has its own variance component. Given that each of the effects is assumed to be independent (and therefore uncorrelated with one another), the expected *total* variance in scores across all Person $\times$ Trait $\times$ Method combinations in the population (in the sample data, these reflect cells in the p $\times$ t $\times$ m data matrix) may be expressed as a simple sum of these variance components:

$$\sigma^2_{\text{expected total}} = \sigma^2_{\text{p}} + \sigma^2_{\text{t}} + \sigma^2_{\text{m}} + \sigma^2_{\text{pt}} + \sigma^2_{\text{pm}} + \sigma^2_{\text{tm}} + \sigma^2_{\text{ptm,r}}. \tag{2}$$

Table 1. Example Multitrait–Multimethod Correlation Matrix Augmented With Expected Correlations Based on Standardized Univariate Generalizability Theory Variance Components

		Method 1			Method 2			Method 3		
	Traits	t1m1	t2m1	t3m1	t1m2	t2m2	t3m2	t1m3	t2m3	t3m3
Method 1	t1m1		$\sigma^2_p+\sigma^2_{pm}$	$\sigma^2_p+\sigma^2_{pm}$	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p	σ^2_p
	t2m1	$r_{t2m1,t1m1}$		$\sigma^2_p+\sigma^2_{pm}$	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p
	t3m1	$r_{t3m1,t1m1}$	$r_{t3m1,t2m1}$		σ^2_p	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$
Method 2	t1m2	$\mathbf{r_{t1m2,t1m1}}$	$\mathrm{r}_{\mathrm{t1m2,t2m1}}$	$\mathrm{r}_{\mathrm{t1m2,t3m1}}$		$\sigma^2_p+\sigma^2_{pm}$	$\sigma^2_p+\sigma^2_{pm}$	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p	σ^2_p
	t2m2	$\mathrm{r}_{\mathrm{t2m2,t1m1}}$	$\mathbf{r_{t2m2,t2m1}}$	$\mathrm{r}_{\mathrm{t2m2,t3m1}}$	$r_{t2m2,t1m2}$		$\sigma^2_p+\sigma^2_{pm}$	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$	σ^2_p
	t3m2	$\mathrm{r}_{\mathrm{t3m2,t1m1}}$	$\mathrm{r}_{\mathrm{t3m2,t2m1}}$	$\mathbf{r_{t3m2,t3m1}}$	$r_{t3m2,t1m2}$	$r_{t3m2,t2m2}$		σ^2_p	σ^2_p	$\boldsymbol{\sigma^2_p+\sigma^2_{pt}}$
Method 3	t1m3	$\mathbf{r_{t1m3,t1m1}}$	$\mathrm{r}_{\mathrm{t1m3,t2m1}}$	$\mathrm{r}_{\mathrm{t1m3,t3m1}}$	$\mathbf{r_{t1m3,t1m2}}$	$\mathrm{r}_{\mathrm{t1m3,t2m2}}$	$\mathrm{r}_{\mathrm{t1m3,t3m2}}$		$\sigma^2_p+\sigma^2_{pm}$	$\sigma^2_p+\sigma^2_{pm}$
	t2m3	$\mathrm{r}_{\mathrm{t2m3,t1m1}}$	$\mathbf{r_{t2m3,t2m1}}$	$\mathrm{r}_{\mathrm{t2m3,t3m1}}$	$\mathrm{r}_{\mathrm{t2m3,t1m2}}$	$\mathbf{r_{t2m3,t2m2}}$	$\mathrm{r}_{\mathrm{t2m3,t3m2}}$	$r_{t2m3,t1m3}$		$\sigma^2_p+\sigma^2_{pm}$
	t3m3	$\mathrm{r}_{\mathrm{t3m3,t1m1}}$	$\mathrm{r}_{\mathrm{t3m3,t2m1}}$	$\mathbf{r_{t3m3,t3m1}}$	$\mathrm{r}_{\mathrm{t3m3,t1m2}}$	$\mathrm{r}_{\mathrm{t3m3,t2m2}}$	$\mathbf{r_{t3m3,t3m2}}$	$r_{t3m3,t1m3}$	$r_{t3m3,t2m3}$	

Note: Cells with italicized text reflect heterotrait–monomethod correlations. Cells with bold text reflect monotrait–heteromethod correlations. Cells with no text formatting reflect heterotrait–heteromethod correlations. σ^2_p = expected proportion of observed variance in trait–method units attributable to person main effects (shared variance that is neither trait nor method specific); σ^2_{pt} = expected proportion of observed variance in trait–method units attributable to Person × Trait interaction effects (shared variance specific to units that share a given trait); σ^2_{pm} = expected proportion of observed variance in trait–method units attributable to Person × Method interaction effects (shared variance specific to units that share a given method).

Typically, in the context of modeling MTMM data, organizational researchers are interested not in expected total variance but, rather, in expected *observed* variance in scores on trait–method units *across* persons. The expected observed variance in scores across persons is the sum of only a subset of the variance components above; namely,

$$\sigma^2_{\text{expected observed}} = \sigma^2_{\text{p}} + \sigma^2_{\text{pt}} + \sigma^2_{\text{pm}} + \sigma^2_{\text{ptm,r}}. \tag{3}$$

Note that trait main effect variance ($\sigma^2{}_{\text{t}}$), method main effect variance ($\sigma^2{}_{\text{m}}$), and Trait × Method interaction effect variance ($\sigma^2{}_{\text{tm}}$) do not contribute to observed variance across persons because the aforementioned effects are constants *across* persons for any given trait–method unit. This is why it is essential that when discussing variance component-based decompositions of MTMM data researchers do not use the terms *trait effects* and *method effects* loosely. In the case of the prototypical fully crossed MTMM design such as this, trait *main* effects and method *main* effects have no impact on observed variance, but other effects involving traits and methods (i.e., ν_{pt} and ν_{pm}) do. As we will note later, the "trait effects" and "method effects" typically discussed in CFA-based decompositions of MTMM data manifest themselves as Person × Trait and Person × Method interaction effects (respectively) in the context of univariate G-theory. Thus, researchers in this area should choose their language carefully when describing and framing trait and method effects.

Estimating Variance Components Underlying MTMM Data

As previously noted, the primary means of estimating variance components in the context of G-theory has historically been ANOVA-based procedures (e.g., Shavelson & Webb, 1991). These procedures required rather arduous manipulations of expected mean squares tables or highly specialized software that was idiosyncratic to G-theory (e.g., GENOVA, urGENOVA; Brennan, 2001). Fortunately, over the past four decades there have been several advances in variance component estimation that can greatly simplify the process (Searle et al., 1992). Procedures for the direct estimation of variance components are now widely available in common statistical packages, and clear examples of the ease with which variance components can be estimated using SAS and SPSS are provided by DeShon (2002) and Putka and McCloy (2008). With regard to MTMM data in particular, Figure 1 illustrates how organizational researchers should structure their raw data to estimate variance components and the SAS or SPSS code necessary for generating estimates for each component in Equation 2.[3] Note that this syntax would remain the same regardless of (a) the number of methods or traits with which one was working and (b) whether or not there were an equal number of traits per method or methods per trait.

In addition to being easy to implement, modern methods of variance component estimation have another key advantage: They can readily deal with missing data and unbalanced designs (DeShon, 1995; Marcoulides, 1990; Putka, Le, McCloy, & Diaz, 2008). This is something that proves difficult for ANOVA-based estimators of variance components characteristic of the G-theory literature. Despite this benefit of modern methods, recent applications of G-theory to MTMM data by organizational researchers have unnecessarily discarded data to achieve a balanced design for purposes of estimating variance components (e.g., Arthur, Woehr, & Maldegen, 2000; Kraiger & Teachout, 1990; Lievens, 2001, 2002). Although modern methods of variance component estimation are clearly advantageous in several respects (see Searle et al., 1992, for a review), the most notable drawback of those methods—largely based on maximum likelihood—is that they can involve rather substantial memory requirements (Bell, 1985). This can make their use prohibitive for larger MTMM data sets.[4] Second, at least as implemented in SAS and SPSS, the aforementioned variance component estimation methods do not readily allow researchers to test any of the assumptions underlying the G-theory model nor to evaluate its fit to a given MTMM data set. Nevertheless, as

person	trait	method	score
1	1	1	5
1	2	1	4
1	3	1	3
1	1	2	4
1	2	2	3
1	3	2	1
1	1	3	2
1	2	3	4
1	3	3	4
2	1	1	3
2	2	1	2
2	3	1	5
2	1	2	1
2	2	2	1
2	3	2	2
2	1	3	4
2	2	3	3
2	3	3	5
.	.	.	.
.	.	.	.
.	.	.	.

SAS Code

```
PROC MIXED METHOD=REML;
CLASS person trait method;
MODEL score = ;
RANDOM person trait method
person*trait person*method trait*method
RUN;
```

SPSS Code

```
VARCOMP
score BY person trait method
/RANDOM = person trait method
/METHOD = REML
/DESIGN = person trait method
person*trait person*method trait*method
/INTERCEPT = INCLUDE.
```

Figure 1. Example code for fitting a univariate generalizability theory model to multitrait–multimethod data
Note: Each row in the data set corresponds to a unique Person × Trait × Method combination.

we highlight below, there is an even simpler approach to estimating variance components that contribute to observed variance across persons that has the benefit of making explicit the link between G-theory variance components and MTMM correlations and that also helps clarify how the G-theory model can be expressed and estimated as a CFA model—which will allow researchers to evaluate the adequacy of its assumptions.

Relations Between Variance Components and MTMM Correlations

What is rarely discussed in the context of applying G-theory to MTMM data in the organizational research literature (at least in the past 25 years) is the link between variance components and elements of an MTMM correlation matrix. As illustrated in Brennan's (2001) recent text on G-theory, and several others previously (e.g., Cronbach et al., 1972; DeShon, 1998; Guilford, 1954; King et al., 1980; Schmitt & Stults, 1986; Stanley, 1961), relationships exist between variance components and the *average* correlations among observed variables or, in the case of MTMM data, among trait-method units. Specifically,

$$\text{Average monotrait} - \text{heteromethod (MTHM) } r = \sigma^2_{\mathrm{p}} + \sigma^2_{\mathrm{pt}}, \tag{4}$$

$$\text{Average heterotrait} - \text{monomethod (HTMM) } r = \sigma^2_{\mathrm{p}} + \sigma^2_{\mathrm{pm}}, \text{ and} \tag{5}$$

$$\text{Average heterotrait} - \text{heteromethod (HTHM) } r = \sigma^2_{\mathrm{p}}. \tag{6}$$

The relationships above reflect the fact that variance components can also be thought of as covariances (Searle et al., 1992). For example, although earlier we described $\sigma^2{}_{\mathrm{p}}$ as variance in trait–method units attributable to person main effects, it can also be defined as shared variance (i.e., covariance) among trait–method units that is neither trait specific nor method specific. Similarly, although we described σ^2_{pt} as variance in trait–method units attributable to Person × Trait interaction effects, it can also be defined as shared variance among trait–method units that is specific to a given trait. Defining variance components in this manner makes apparent the relationships established in Equations 4 through 6. For example, based on the univariate G-theory model presented in Equation 1, the expected correlation between two different measures of a common trait (i.e., MTHM *r*) is composed of two sources of covariation—one source is general (i.e., σ^2_{p}), and the other is trait specific (i.e., σ^2_{pt}). Similarly, the expected correlation between two different traits measured by a common method (i.e., HTMM *r*) is composed of two sources of covariation—again, one source is general (i.e., σ^2_{p}), but in this case, the other is method specific (i.e., σ^2_{pm}).

By rearranging the formulas above, it is a simple matter to generate *standardized* variance components as a function of average MTMM correlations, namely,[5]

$$\sigma^2_{p} = \text{Average HTHM } r, \tag{7}$$

$$\sigma^2_{pt} = \text{Average MTHM } r - \text{Average HTHM } r, \tag{8}$$

$$\sigma^2_{pm} = \text{Average HTMM } r - \text{Average HTHM } r, \tag{9}$$

$$\sigma^2_{ptm,r} = 1 - \sigma^2_{p} + \sigma^2_{pt} + \sigma^2_{pm}. \tag{10}$$

Linking Variance Components to Notions of Convergent–Discriminant Validity

Given the relation between MTMM correlations and variance components, clear links can be made between G-theory variance components and Campbell and Fiske's (1959) notions of convergent and discriminant validity. For example, past discussions have equated the magnitude of person main effect variance (σ^2_{p}) with convergent validity (e.g., Dickinson, 1986). However, if one views convergent validity as a reflection of sources of consistency among different measures of a given trait, the Person × Trait interaction variance (i.e., σ^2_{pt}) will also contribute to such consistency because it is a source of common variance for measures that share a trait in common. Other researchers have defined convergent validity in terms of the magnitude of the Person × Method variance component (i.e., σ^2_{pm}; Kane, 1996; Kraiger & Teachout, 1990; Lievens, 2001). However, equating the magnitude of such components to evaluations of convergent validity is problematic in that the size of σ^2_{pm} does not index the extent of consistency (i.e., convergence) among different measures of a common trait. Similarly, past research has equated discriminant validity with the magnitude of Person × Trait interaction variance (i.e., σ^2_{pt}; Arthur et al., 2000; Dickinson, 1986; Lievens, 2001), however, as we present in Table 2, Campbell and Fiske (1959) actually describe three conditions that should be established for discriminant validity. Thus, upon revisiting the extant literature, one is left with a fundamental question, namely, How do the variance components stemming from G-theory relate to the traditional notions of construct-related validity? Table 2 provides clarification of these issues that directly follow from relations established in Equations 4 through 10 and descriptions of criteria for convergence and discrimination offered by Campbell and Fiske. Furthermore, Table 2 offers G-theory-based indices for evaluating overall evidence for convergent and discriminant validity within the MTMM system.

Table 2 shows that from a univariate G-theory perspective, evidence of convergent validity for traits assessed in the MTMM system is reflected by the magnitude of the sum of σ^2_{p} and σ^2_{pt}, which reflects the proportion of expected observed variance in trait–method units attributable to (a) person main effects and (b) shared variance among units that is specific to a given trait. Unlike evidence of

Table 2. Univariate Generalizability Theory (G-theory) Analogues of Campbell and Fiske's (1959) Definitions of Convergence, Discrimination, and Method Effects

Criterion	Campbell and Fiske's Definition	Univariate G-Theory Analogue	Univariate G-Theory Type Index	Interpretation
Convergence	Entries in the validity diagonal (monotrait–heteromethod *r*s) should be different from zero and sufficiently large to encourage further examination of validity	The sum of σ^2_p and σ^2_{pt} should be greater than zero and sufficiently large	**C1** $= \boldsymbol{\sigma^2_p + \sigma^2_{pt}} =$ (*Ave* HTHM *r*) + (*Ave* MTHM *r* – *Ave* HTHM *r*) = *Ave* MTHM *r*	C1 = The proportion of expected observed variance in trait–method units attributable to (a) person main effects and (b) shared variance among units that is specific to a given trait
Discrimination (Condition 1)	Entries in the validity diagonal (monotrait–heteromethod *r*s) should be higher than values in their corresponding columns and rows in the heterotrait–heteromethod triangles	$\sigma^2_p + \sigma^2_{pt}$ should be larger than σ^2_p; i.e., σ^2_{pt} should be greater than zero	**D1** $= \boldsymbol{\sigma^2_{pt}} =$ *Ave* MTHM *r* – *Ave* HTHM *r*	D1 = The proportion of expected observed variance in trait–method units attributable to shared variance among units that is specific to a given trait
Discrimination (Condition 2)	Entries in the validity diagonal (monotrait–heteromethod *r*s) should be higher than values in the heterotrait–monomethod triangles	$\sigma^2_p + \sigma^2_{pt}$ should be larger than $\sigma^2_p + \sigma^2_{pm}$; i.e., σ^2_{pt} should be larger than σ^2_{pm}	**D2** $= \boldsymbol{\sigma^2_{pt} - \sigma^2_{pm}} =$ (*Ave* MTHM *r* – *Ave* HTHM *r*) – (*Ave* HTMM *r* – *Ave* HTHM *r*) = *Ave* MTHM *r* – *Ave* HTMM *r*	D2 = The difference in proportions of expected observed variance accounted for by trait-specific vs. method-specific shared variance
Discrimination (Condition 3)	The same pattern of trait interrelationships is shown in all of the heterotrait triangles, in both the monomethod and heteromethod blocks	No direct analogue, as all trait interrelations are expected to reflect $\sigma^2_p + \sigma^2_{pm}$ in monomethod blocks, and all trait intercorrelations are expected to reflect σ^2_p in heteromethod blocks	An option here is to calculate local standardized root mean square residual–like indices for trait intercorrelations within (a) each monomethod block to assess whether the G-theory implied structure fits equally as well for each method and (b) each heteromethod block to assess whether the G-theory implied structure fits equally as well for each pair of methods (i.e., m1 vs. m2, m1 vs. m3, m2 vs. m3)	
Method variance	Indicated by the difference in level of correlation between the parallel values of the monomethod block and the heteromethod blocks	$\sigma^2_p + \sigma^2_{pm}$ is larger than σ^2_p; i.e., σ^2_{pm} is greater than zero	**MV** $= \boldsymbol{\sigma^2_{pm}} =$ *Ave* HTMM *r* – *Ave* HTHM *r*	MV = The proportion of expected observed variance in trait–method units attributable to shared variance among units that is specific to a given method

Note: The interpretations above assume one is dealing with variance components standardized against observed variance (i.e., $\sigma^2_p + \sigma^2_{pt} + \sigma^2_{pm} + \sigma^2_{ptm,r} = 1$).

convergent validity, garnering evidence of discriminant validity for traits assessed in the MTMM system is more complex and is a function of three conditions. Condition 1 is reflected by the magnitude of σ^2_{pt}, which reflects the proportion of expected observed variance in trait–method units attributable to shared variance among units that is specific to a given trait. Condition 2 is reflected by the magnitude of the difference between σ^2_{pt} and σ^2_{pm}, which reflects the difference in proportions of expected observed variance accounted for by trait–specific versus method–specific shared variance. Finally, Condition 3 cannot readily be evaluated by the standard output associated with a univariate G-theory decomposition of variance. Indeed, the univariate G-theory model implicitly assumes that this condition will be met in that it specifies that all trait interrelations are expected to reflect $\sigma^2_p + \sigma^2_{pm}$ in monomethod blocks and that all trait intercorrelations are expected to reflect σ^2_p in heteromethod blocks. Thus, under the univariate G-theory model, Condition 3 is assumed to be satisfied, but as we will note below, this assumption, as well as others that underlie the univariate model, can be readily tested.

The Structural Model Implied by the Univariate G-Theory Model of MTMM Data

As the upper diagonal of Table 1 reveals, the univariate G-theory model implies a very simple correlational structure: (a) All HTHM correlations are expected to equal σ^2_p, (b) all MTHM correlations are expected to equal $\sigma^2_p + \sigma^2_{pt}$, and (c) all HTMM correlations are expected to equal $\sigma^2_p + \sigma^2_{pm}$. Furthermore, given the assumptions underlying the univariate G-theory model, one can also frame it in terms of a highly constrained CFA model, as depicted in Figure 2.

Based on the model in Figure 2, one can readily see that the observed variance in any given trait–method unit will be decomposed into the four components—σ^2_p, σ^2_{pt}, $\sigma^2_{pm,}$ and $\sigma^2_{ptm,r}$ (per Equation 3 presented earlier)—and that the CFA framework offers an alternative means to estimate univariate G-theory variance components (Marcoulides, 1996; Raykov & Marcoulides, 2006). Clearly, the univariate G-theory model implies a very constrained model whose only unknown parameters can simply be estimated from average correlations. Nevertheless, fitting the G-theory model within a CFA framework offers several advantages: (a) It will produce indices of model fit that can be used to compare the univariate G-theory model to other MTMM CFA models; (b) it gives MTMM researchers a meaningful starting point for exploring the latent structure of MTMM data (the model is simple enough that it can always be fitted to data without fear of convergence or admissibility issues); and (c) it provides researchers with a method for easily testing assumptions underlying the G-theory model—something that is not possible using traditional variance component estimation procedures.[6] For example, using the model depicted in Figure 2 as a basis, researchers can easily relax constraints to achieve a more refined decomposition of trait–method unit variance and test various assumptions behind the univariate G-theory model such as zero covariance among different Person × Trait effects and different Person × Method effects, respectively. We revisit issues of how the univariate G-theory model can be relaxed in later sections.

The Multivariate G-Theory Model and MTMM Data

Given all of the constraints it imposes, one may argue that the univariate G-theory model provides an overly simplified decomposition of the variance in each trait–method unit. Indeed, some might argue that the univariate G-theory model, which is typically applied to assess the generalizability of scores from a measure of a single trait or construct, may be suboptimal for analyzing MTMM data given that multiple traits are being assessed. For example, the univariate G-theory model (as applied here) implies that there will be no common variance between different traits (measured by different methods) beyond that accounted for by a general person main effect, something that may be unlikely to hold in many MTMM studies.

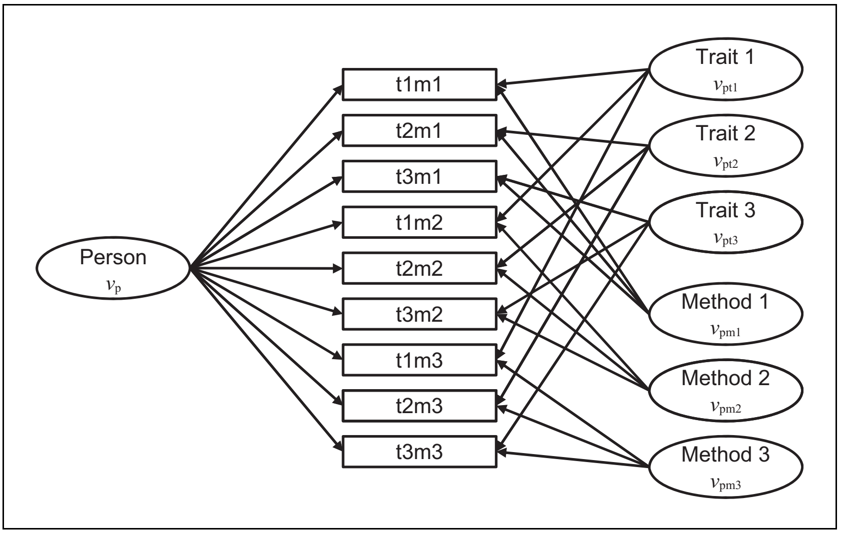

Figure 2. Univariate generalizability theory model of multitrait–multimethod data expressed as a structural model

Note: Covariances (i.e., ϕs) and error variances (i.e., δs) are omitted for the sake of parsimony. Constraints imposed on this model are as follows: (a) equal trait factor variances (i.e., $\sigma^2_{pt} = \phi_{pt1} = \phi_{pt2} = \phi_{pt3}$), (b) equal method factor variances (i.e., $\sigma^2_{pm} = \phi_{pm1} = \phi_{pm2} = \phi_{pm3}$), (c) equal unique variances (i.e., $\sigma^2_{ptm,r} = \delta_{t1m1} = \delta_{t1m2} = \delta_{t1m3} = \delta_{t2m1} = \delta_{t2m2} \ldots = \delta_{t3m3}$), (d) all factor covariances (off diagonals in the Θ_δ matrix) are constrained to zero, (e) all error covariances (off diagonal in the Θ_δ matrix) are constrained to zero, and (f) all factor loadings (λs) are constrained to one.

Fortunately, the univariate G-theory model can easily be adjusted when fitted within a CFA framework not only to systematically evaluate the constraints that the model imposes on the data (as described above) but also to evaluate whether a multivariate G-theory model may be more appropriate for one's data. Although rarely discussed in the context of organizational research literature, one can view fitting a multivariate G-theory model to MTMM data as akin to fitting a univariate G-theory model for each trait separately, while accounting for the fact that components of observed variance for each trait–method unit will be correlated due to covariances among latent traits and covariances among trait–method units that share a method in common.

To illustrate these concepts, Figure 3 shows the multivariate G-theory model expressed in terms of a CFA framework. This model essentially amounts to fitting a Person $\times$ Method univariate G-theory model for each trait separately, yet it allows for components of observed variance in the p $\times$ m design underlying each trait, namely, $\sigma^2_{p,pt}$ and $\sigma^2_{pm,r}$, to differ and be correlated across traits.[7] These correlations reflect the fact that the same methods are used to assess each person on each trait. From a CFA perspective, this amounts to fitting a constrained version of the CT-CU model (Scullen, 1999) in which uniquenesses for trait–method units that share a method in common are allowed to correlate, whereas correlations among uniquenesses for trait–method units that do not share a method are constrained to zero. The full set of constraints implied by the multivariate G-theory model is provided in the note accompanying Figure 3.

Based on the structure depicted in Figure 3, the multivariate G-theory model produces estimates of 12 parameters: (a) three estimates for $\sigma^2_{p,pt}$ (a separate estimate for each trait); (b) three estimates for $\sigma^2_{pm,r}$ (again, a separate estimate for each trait); (c) three estimates for $\sigma_{p,pti,p,ptj}$ (a separate estimate for each pair of traits), and (d) three estimates for $\sigma_{pm,ri,pm,rj}$ (again, a separate estimate for each pair of traits).[8] Definitions for each of these components are presented in Table 3.

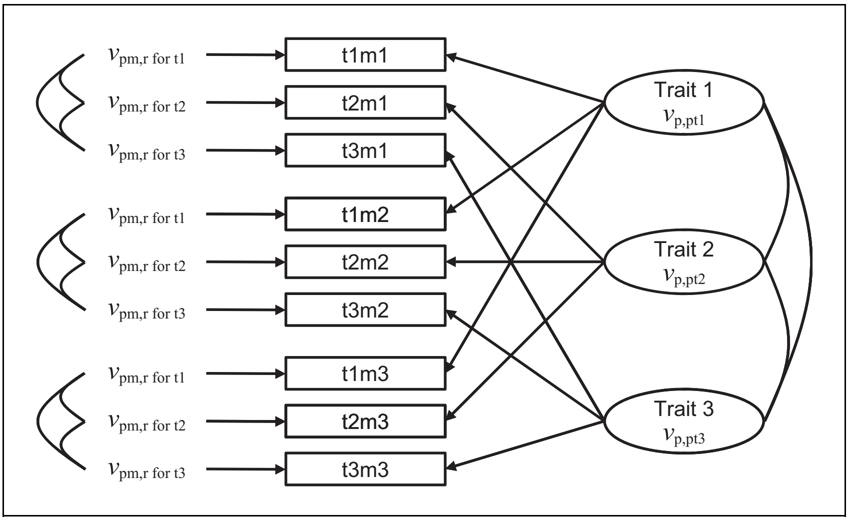

Figure 3. Multivariate generalizability theory model of multitrait–multimethod data expressed as a structural model

Note: Constraints imposed by the model are as follows: (a) equal unique variances for trait-method units involving a given trait (i.e., $\sigma^2_{pm,r \text{ for Trait } i} = \delta_{tim1} = \delta_{tim2} = \delta_{tim3}$), resulting in estimation of three unique variances—one for each trait; (2) equal error covariances for pairs of trait–method units (trait *i,* trait *i′*) that share a method in common (i.e., $\sigma_{pm,r \text{ for trait } i,\ pm,r \text{ for trait } i'} = \delta_{tim1,ti'm1} = \delta_{tim2,ti'm2} = \delta_{tim3,ti'm3}$), resulting in estimation of three error covariances—one for each pair of traits; (3) error covariances for pairs of trait–method units that do not share a method in common are constrained to zero; and (4) all factor loadings (λs) are constrained to one.

It is important to note that the substantive meaning of person main effects (as well as their variance components) changes relative to the univariate G-theory model. Recall that in the univariate model, person main effects conceptually reflected the expected value of a person's score across the population of traits *and* methods. Here, they conceptually reflect the expected value of a person's score on the *given trait* across the population of methods *only*. Furthermore, recall that under the univariate G-theory model, variance attributable to person main effects could be viewed as shared variance (covariance) among trait–method units that was neither trait specific nor method specific. In the multivariate G-theory model, person main effects now reflect shared variance among trait–method units that share a trait, but this shared variance may not be simply trait specific; it could reflect variance that is shared across traits (i.e., more general in nature, such as the person main effect in the univariate G-theory model) but still unrelated to method.

Relations Between Multivariate G-Theory Model Parameters and MTMM Correlations

As was the case with univariate G-theory, there is an explicit link between the variance and covariance components that underlie multivariate G-theory and elements of an MTMM correlation matrix. More generally, Cronbach et al. (1972) noted that relationships exist between variance and covariance components and the *average* variances and covariances among observed variables. In the case of MTMM data, these observed variables represent trait–method units. Such relationships allow researchers to estimate the variance and covariance components underlying the multivariate G-theory model without the use of idiosyncratic G-theory software (e.g., mGENOVA; Brennan, 2001) or even SEM software. Specifically, Table 3 shows how variance and covariance components underlying the multivariate G-theory model relate to (and indeed can be estimated from) observed variances and covariances from an MTMM matrix.

Table 3. Links Between Multivariate Generalizability Theory Model Parameters and Observed Variances and Covariances

Observed Variances and Covariances	
σ^2_{xti}	Average of variances (MTMM diagonals) for trait–method units involving t_i = expected observed variance for any given trait *i* method unit (xt_i)
$\sigma_{xti,xti'}$	Average of HTMM covariances between trait *i* and trait *i*′ method units = expected observed covariance for trait *i* and trait *i*′ method units that share a method in common
Variance Component Estimates	
$\sigma^2_{p,pti}$	Average of MTHM covariances for trait–method units involving t_i = person main effect variance + Person × Trait interaction variance for trait–method units involving t_i
$\sigma^2_{pm,ri}$	Difference: $\sigma^2_{xti} - \sigma^2_{p,pti}$ = Person × Method interaction variance + residual variance for trait–method units involving t_i
Covariance Component Estimates	
$\sigma_{p,pti,p,pti'}$	Average of HTHM covariances between trait *i* method units and trait *i*′ method units = covariance between person main effects for traits *i* and *i*′
$\sigma_{pm,ri,pm,ri'}$	Difference: $\sigma_{xti,xti'} - \sigma_{pti,pti'}$ = covariance between residual effects for trait *i* and trait *i*′ method units that share a method in common

Note: MTMM = monotrait–monomethod; HTMM = heterotrait–monomethod; MTHM = monotrait–heteromethod; HTHM = heterotrait–heteromethod.

Contrasting Univariate and Multivariate G-Theory Perspectives on MTMM Data

Since the multivariate G-theory model presents researchers with more complexity (and thus more information) than the univariate G-theory model, there are several reasons organizational researchers may initially be inclined to find the multivariate model more valuable for examining MTMM data. First, the multivariate G-theory model allows for estimation of (a) *unique* sets of variance components for each trait and (b) correlations among the latent components of observed score variance—neither of which can be estimated by the univariate G-theory model. Indeed, the general nature of variance components produced by the univariate G-theory model somewhat limits its utility with regard to informing judgments regarding evidence of convergent and discriminant validity for *specific* measures. For example, recall that earlier we showed how variance components from the univariate G-theory model allows one to make overall judgments with regard to convergent–discriminant validity for the *system* of traits and methods as a whole (per Figure 2)—but not judgments with regard to specific traits or methods—which is often of interest when one attempts to establish validity evidence for scores produced by a given measure (Kane, 2002; Putka & Sackett, 2010). Since unique sets of variance components and covariance components are estimated for each trait under the multivariate G-theory model discussed here, it can be viewed as providing a more refined method for exploring convergent validity evidence for individual measures within an MTMM system than can be offered by the univariate G-theory model. Nevertheless, for reasons we note below, the multivariate G-theory framework suffers from limitations when evaluating evidence of discriminant validity.

Second, even though the multivariate G-theory model affords researchers more flexibility than the univariate G-theory model, it still implies that a very parsimonious latent structure underlies MTMM correlations (i.e., it is unlikely to lead to the convergence and admissibility problems that plague more traditional, *df*-hungry, correlated trait–correlated method CFA models). Indeed, much like the univariate G-theory model, organizational researchers can view the multivariate G-theory model as a convenient (in terms of convergence and admissibility issues) point of departure for modeling MTMM data. Furthermore, much like the univariate G-theory model, it would also be relatively straightforward to systematically relax constraints of the multivariate G-theory model to assess the veracity of its assumptions and explore whether a less parsimonious model may provide a better fit to one's data.

Despite its advantages, however, the multivariate G-theory model suffers from a few key limitations that have yet to be recognized in the literature. As noted earlier, and made apparent in Figure 3, it is analogous in structure to CT-CU models, which have been heavily criticized in the organizational literature (e.g., Lance et al., 2002). Just like the CT-CU model, this model assumes that no correlations exist among method effects. To the extent this assumption is violated (*and* method effects are nonzero), estimates of trait loadings and trait correlations in CT-CU models will be biased upward. In the context of a multivariate G-theory model applied to MTMM data, such unaccounted for method correlations would manifest themselves as upwardly biased estimates of $\sigma^2_{\mathrm{p,pt}}$ (confounded person main effect and Person × Trait interaction effect variance components) and $\sigma_{\mathrm{p,pt}i,\mathrm{p,pt}j'}$ (covariance components among the aforementioned effects), or at the very least require that researchers take into account the potential presence of such correlated effects when interpreting these estimates. Given this limitation, we suggest that when adopting a G-theory-based modeling approach, researchers pay particular attention to modification indices that suggest the presence of unaccounted for covariance among residuals. Although the focus of the criticism above is on the multivariate G-theory model, as we note below the issue of unaccounted for covariance among score effects can also be an issue for the univariate G-theory model and should be considered as part of a more in-depth modeling process.

Another potential drawback of multivariate G-theory model is that it fails to result in a partitioning of trait–method unit variance into distinguishable trait-specific, method-specific, *general,* and residual variance components (as the univariate G-theory model does). As noted above, within the multivariate G-theory model, trait-specific variance and general factor variance (person main effect variance in the context of the univariate G-theory model) are confounded, and method-specific and residual variance are confounded (recall Note 7). This confounding creates problems for researchers wishing to evaluate discriminant validity evidence based on multivariate G-theory results.

For example, consider the criteria for the first two conditions for discriminant validity evidence shown in Table 2. Under univariate G-theory, unique estimates of σ^2_{pt} and σ^2_{pm} are provided, thus allowing one to evaluate the first two conditions for discriminant validity. Under the multivariate G-theory model, σ^2_{pt} is confounded with σ^2_{p} (a source of convergent validity, but not discriminant validity) under the $\sigma^2_{\mathrm{p,pt}}$ variance components, thus making it difficult to evaluate evidence for discriminant validity (Condition 1); similarly, σ^2_{pm} is confounded with $\sigma^2_{\mathrm{ptm,r}}$ (a source of residual error) under the $\sigma^2_{\mathrm{pm,r}}$ variance components, thus making it difficult to evaluate evidence for discriminant validity (Condition 2). Thus, the confounding of effects in the multivariate G-theory model clearly creates difficulty for evaluating evidence of discriminant validity relative to the univariate G-theory model where key effects are not confounded.

A G-Theory-Based Process for Modeling MTMM Data

Whether researchers start their modeling process with a univariate or multivariate G-theory model should be dictated by their purpose and expectation regarding the functioning of the MTMM system. For example, if the researcher is concerned with isolating the influence of person main effects on variance in trait–method units (e.g., wishes to partition trait–method unit variance into trait-specific, method-specific, residual, and general components), then the univariate G-theory model would provide such information. If the researcher is not as interested in the influence of a general person factor and more interested in obtaining trait-specific variance component estimates, then the multivariate G-theory model may provide a better starting point. Of course there is nothing preventing the researcher from fitting both the univariate and multivariate G-theory model, as each provides a somewhat different perspective on the functioning of the MTMM system.

With regard to where the researcher goes from the basic G-theory models, it would again depend on the researcher's purpose for modeling the data. For example, with either model the researcher may be interested in a decomposition that is more tailored to specific trait–method units. If that is the case, the researcher can achieve a univariate G-theory-style decomposition of variance for each trait–method unit separately by (a) fixing the general (person), trait, and method factor variances to one, (b) allowing the *loadings* in Figure 2 to be freely estimated, and (c) freeing the equality constraints on the uniquenesses. With the resulting parameter estimates, the researcher can calculate unique general, trait-specific, method-specific, and residual variance components for each trait–method unit. In this case, the squared loadings associated with general, trait, and method factors would reflect variance components for each trait–method unit. Similarly, within the multivariate G-theory model, such a tailored decomposition could be achieved by (a) fixing the trait factor variances to one, (b) allowing the loadings in Figure 3 to be freely estimated, and (c) freeing the equality constraints on the uniquenesses and covariances for uniqueness. However, such fully unconstrained models are highly similar to traditional CFA models of MTMM data and thus are likely to result in model convergence and inadmissibility problems. That is, as more constraints are removed from the model (i.e., more parameters are freely estimated), the model will be less likely to converge on admissible solutions. A potential solution is to successively relax model constraints in order to obtain the maximum amount of information while still allowing appropriate estimation.

Regardless of whether the researcher is interested in the traditional, highly constrained univariate and multivariate G-theory models or the less constrained versions that result from freeing up equality constraints, a critical subsequent step in the process would be to evaluate G-theory model assumptions that have direct implications for the accuracy (or at the very least, interpretation) of the variance and covariance component estimates they produce. For example, the univariate G-theory literature has suggested that unaccounted for covariances among residual effects would lead to inflated estimates of person main effect variance and to underestimates of residual variance (e.g., Bost, 1995; Smith & Luecht, 1992). The effects predicted by the work of these researchers, coupled with Box's (1954) earlier work, would also suggest that unaccounted for covariance among other sources of observed variance (e.g., Person $\times$ Trait and Person $\times$ Method interaction effects) would lead to inflated estimates of person main effects in the context of applying a univariate G-theory model to MTMM data because it would be the person main effect term that is relied upon to account for such covariance. Nevertheless, Smith and Luecht (1992) noted that "if the correlation between all pairs of effects are fairly homogeneous or vary in some random fashion, it can be argued that they are inherent to the measurement context or domain, and are therefore correctly accommodated in generalizability study variance components" (p. 230). What this means is that the additional covariance that is being reflected in the person main effect does not reflect "bias" in the parameter estimate per se but, rather, *replicable* between-person variance that would manifest itself if another sample was drawn from the population. Thus, we advise researchers who utilize the univariate G-theory model to examine model residuals to carefully evaluate whether additional covariance may be unaccounted for among factors and uniqueness specified by the model.

Although there has been work done in the univariate G-theory literature regarding violating the assumption of uncorrelated effects, little has been done in the multivariate G-theory literature in this regard. It is likely that this has been the case because multivariate G-theory represents a framework that was specifically designed to account for substantively meaningful correlations among score effects that arise from the nature of the traits studied and measurement design. Nevertheless, given the comparability between the CT-CU model and the multivariate G-theory model, it is clear that there are still issues of unaccounted for covariance that may bias (or again, at the very least, change the interpretation of) multivariate G-theory variance and covariance parameter estimates. Thus, when applying G-theory methods to MTMM data, a key advantage of applying them within a CFA framework (i.e., fitting them using SEM software) is that the constraints above can be tested to see if they hold.

A Word of Caution on Judging the Overall Fit of G-Theory Models

Historically, applications of G-theory have not been overly concerned with notions of model fit nor with testing the assumptions underlying G-theory models. We believe that this is due in part to the fact that G-theory models have been framed not from a CFA perspective but, rather, from an ANOVA perspective. We also believe this reflects the fact that unlike most CFA models applied to MTMM data, G-theory is focused not on replicating the whole covariance structure at the level of individual trait–method units but, rather, on what may best be thought of as the average covariance structure among trait–method units where the nature of the averages are dictated by one's measurement design. Regardless of the G-theory literature's lack of attention to these issues, if G-theory methods are to play a more prominent role in evaluating MTMM data, it is critical to address the issues of model fit since they are central to a model-fitting paradigm.

Relative to most CFA models examined in the MTMM literature, G-theory models and, in particular, the univariate G-theory model, are highly constrained, thus decreasing the likelihood of finding high levels of absolute fit between the G-theory models and the observed data. Often, the purpose of modeling MTMM matrices with CFA methods is to specify and identify a theoretically meaningful structural model that reproduces the *entire* matrix as accurately as possible. As such, common fit indices, whether designed to assess fit in an absolute sense (e.g., standardized root mean square residual [SRMR], chi-square) or to assess fit using a metric that considers model parsimony (e.g., CFI, RMSEA), are based at some level on deviations between observed and model-implied variances and covariances among all individual variables in the matrix.

As noted above, when considering the use of constrained G-theory methods for modeling MTMM data, it is critical to recognize that such methods attempt to reproduce an *average* covariance structure, not the specific covariance structure among each of the individual variables. In the case of applying G-theory methods to MTMM data, the averages are taken across related sets of trait–method units (e.g., trait–method units that share a trait in common, trait–method units that share a method in comment). As such, G-theory methods do not attempt to reproduce heterogeneity in variances–covariances that may exist within those related sets (i.e., as reflected in the full trait–method unit covariance matrix). From a model evaluation perspective, this point is very important to recognize in that it can lead to situations in which a univariate or multivariate G-theory model can provide poor fit to the data based on commonly suggested fit indexes, but the parameter estimates they produce are accurate and representative of the data.

For example, consider a situation in which a researcher fits a univariate G-theory model and the assumptions underlying the model hold with the exception that variances for Person × Trait, Person × Method, and residual effects (i.e., σ^2_{pt1}, σ^2_{pt2}, σ^2_{pt3}; σ^2_{pm1}, σ^2_{pm2}, σ^2_{pm3}, $\sigma^2_{pt1m1,r} \ldots \sigma^2_{pt3m3,r}$) vary, but their averages equal the variance components implied by the univariate G-theory model (i.e., σ^2_{pt}, σ^2_{pm}, and $\sigma^2_{ptm,r}$, respectively). In this example, although relaxing the equality constraints imposed on these variances by the univariate G-theory model may improve model fit–based commonly used fit indexes, relaxing such constraints makes no difference when it comes to the accuracy of the parameter estimates of interest in the univariate G-theory model (i.e., σ^2_{pt}, σ^2_{pm}, $\sigma^2_{ptm,r}$). This logic is also apparent in our earlier discussion that shows how variance and covariance components underlying the univariate and multivariate G-theory models can easily be estimated by taking averages of observed covariances. Whether or not there is heterogeneity in those covariances across the trait–method units on which they are calculated (which the commonly used fit indices would be sensitive to) does not matter so much from a parameter estimation perspective because the average of those covariances remains the same (Marcoulides, 1996).

Although the discussion above appears to suggest that judging the quality of information provided by the univariate G-theory model based on commonly used fit indices can be misleading, this does not mean that *model fit* is irrelevant in the context of G-theory. For example, although the equality constraints imposed by G-theory may have few implications for estimating average variance and covariance components, as we noted earlier, if other constraints fail to hold, namely, uncorrelated effects, biased estimates can result. The problem with using indices of overall model fit to deduce such problems is that they cannot pinpoint whether the source of misfit in the model is due to factors that would bias univariate or multivariate G-theory parameter estimates or factors that would not bias such estimates but would (if accounted for) enable the researcher to obtained a more refined decomposition of trait–method unit variance. Given the factors above, we would urge caution on the part of researchers who use overall fit indices to judge the quality of the univariate and multivariate G-theory model and instead encourage them to focus more attention on evaluating constraints that would result in bias, or at least change the interpretation of, their parameter estimates of interest.

Empirical Illustration of a G-Theory-Based Modeling Process for MTMM Data

To make the application of the G-theory methods described above more tangible, we applied them to a real MTMM data set. The data were drawn from a sample of 347 participants completing measures as part of a promotion system in a large public-sector organization. Given that the analyses performed here are simply for illustration, we do not elaborate on the nature of the traits or methods examined. The measurement design examined here involved three methods and three traits, with each trait being assessed by each method (fully crossed design). This design allowed us to create a nine-by-nine covariance matrix among nine trait–method unit scores (see Table 4).

Prior to running the G-theory analyses, we analyzed the data presented in Table 4 using a traditional parameterization of an MTMM CFA model, that is a 3 correlated trait–3 correlated method (3CT-3CM) model. This model failed to converge to an admissible solution. Similar convergence problems were encountered for a single-trait–3 correlated method (1T-3CM) model. These convergence issues are typical with the modeling of MTMM data and highlight the potential value in using a G-theory-based approach.

Univariate G-Theory Analyses

We first analyzed the data in Table 4 with the univariate G-theory approach described above. Our primary objective is to demonstrate the use of the CFA-based approach to the estimation of the relevant variance components. We also seek to demonstrate the equivalence of this approach with more traditional G-theory computational methods. Thus, we utilized average MTMM correlations as input to Equations 7 through 10, the VARCOMP procedure in SPSS, and LISREL to estimate univariate G-theory model variance components. For the SPSS analyses, we utilized the syntax provided in Figure 1. As previously discussed, this method provides estimates not only for components of *observed* variance in trait–method units (i.e., σ^2_{p}, σ^2_{pt}, σ^2_{pm}, and $\sigma^2_{ptm,r}$) but also for those components of *total* variance that are not captured by the former methods (i.e., σ^2_{t}, σ^2_{m}, and σ^2_{tm}). For the CFA-based analysis, we used the LISREL syntax provided in the appendix. This approach utilizes the MTMM covariance matrix as the input data (as opposed to the raw data in the SPSS analyses) and thus provides estimates only for the components that are a function of the observed variance (i.e., σ^2_{p}, σ^2_{pt}, σ^2_{pm}, and $\sigma^2_{ptm,r}$).

Results of these analyses are presented in Table 5. The three columns in Table 5 indicate that estimating variance components using average MTMM correlations, the SPSS VARCOMP procedure, and CFA resulted in the same decomposition of observed variance. Specifically, these analyses revealed that the Person $\times$ Method interaction effect was the largest component of observed

Table 4. Multitrait–Multimethod Correlation and Covariance Matrices for Empirical Demonstration

	Traits	Method 1 t1m1	Method 1 t2m1	Method 1 t3m1	Method 2 t1m2	Method 2 t2m2	Method 2 t3m2	Method 3 t1m3	Method 3 t2m3	Method 3 t3m3
Method 1	t1m1		.541	.520	.121	.187	.225	.253	.196	.204
	t2m1	*.675*		.535	.100	.158	.153	.182	.174	.174
	t3m1	*.671*	*.705*		.104	.149	.220	.168	.156	.198
Method 2	t1m2	**.148**	.125	.134		.484	.430	.135	.132	.092
	t2m2	.234	**.203**	.198	*.608*		.435	.128	.171	.094
	t3m2	.292	.203	**.302**	*.559*	*.581*		.168	.189	.106
Method 3	t1m3	**.258**	.190	.181	**.138**	.134	.182		.822	.455
	t2m3	.220	**.200**	.185	.148	**.197**	.226	*.771*		.443
	t3m3	.259	.266	**.266**	.117	.123	**.144**	*.482*	*.518*	

Note: Correlations in italicized text reflect heterotrait–monomethod correlations. Correlations in bold text reflect monotrait–heteromethod correlations. Correlations with no text formatting reflect heterotrait–heteromethod correlations. Correlations are below the diagonal, and covariances are above the diagonal.

variance (σ^2_{pm} = .363; 43.2% of observed variance), followed by the Person × Trait × Method interaction (which is completely confounded with residual error; $\sigma^2_{ptm,r}$ = .306; 36.4% of observed variance), the person main effect (σ^2_p = .156; 18.6% of observed variance), and the Person × Trait interaction (σ^2_{pt} = .015; 1.8% of observed variance).

As described earlier, we can interpret the variance components reported in Table 5 from the perspective of Campbell and Fiske's (1959) definitions of convergent and discriminant validity (per Table 2). For these data, the convergence index (C1) is moderate (.171; i.e., 17.1% of observed variance is due to person main effects or shared variance specific to a given trait), and both discrimination indices suggest poor discrimination (D1 = .015 and D2 = –.348). Specifically, the first discrimination index indicates that only 1.5% of observed variance in trait–method units is due to trait-specific variance. Contrasting this value with the convergent validity index (C1) reveals that the majority of convergence is attributable to general person main effects. Furthermore, the second discrimination index indicates that the percentage of shared variance specific to a given trait (i.e., 1.5%) is 34.8 percentage points *lower* than the percentage of shared variance specific to a given method (i.e., 36.3%)—the latter reflecting what is labeled as *method variance* in the Campbell and Fiske framework. Overall, the univariate analysis suggest that measures within this MTMM system demonstrate poor construct-related validity (on average) in terms of both convergence and discrimination.

Although all three computational approaches converge on similar variance component estimates, two benefits of fitting the univariate G-theory model within a CFA framework is that it (a) provides flexibility for relaxing equality constraints (for those interested in a more detailed decomposition) and (b) offers methods for identifying potential sources of misspecification. When framed from a CFA perspective, LISREL model fit statistics indicate that the univariate G-theory model provides an adequate fit of the model to the data, at best (χ^2 = 182.078, $p < .001$; RMSEA = .115; NNFI = .921; CFI = .910; SRMR = .102; goodness-of-fit index [GFI] = .871); however, as we noted earlier, such statistics arguably have little meaning given the focus of the univariate G-theory model on reproducing the average covariance structure, not the structure at the level of individual trait–method units.

Finally, to illustrate how an examination of model residuals can inform the evaluation of univariate G-theory model solutions, we return to the LISREL results for the univariate G-theory model, paying particular attention to the potential correlations among trait and among method factors.

Table 5. Univariate Generalizability Theory Model Variance Component Estimates

	MTMM[a]	SPSS VARCOMP	LISREL
σ^2_p	.190 (190)	.156 (.186)	.157 (.187)
σ^2_t	–	.051	–
σ^2_m	–	.149	–
σ^2_{pt}	.016 (.016)	.015 (.017)	.014 (.016)
σ^2_{pm}	.429 (.429)	.363 (.432)	.361 (.430)
σ^2_{tm}	–	.076	–
$\sigma^2_{ptm,r}$	.365 (.365)	.306 (.364)	.307 (.366)

Note: Raw variance components are shown outside of parentheses; proportions of observed variance associated with components that contribute to observed variance are shown in parentheses. Also note, raw variance components for the multitrait–multimethod (MTMM) approach are based on using standardized statistics as input (MTMM correlations), whereas the SPSS and LISREL analyses are not; thus, only the proportions of observed variance match across approaches, not the raw variance component estimates.
[a]These calculations are based on using average MTMM correlations from Table 4 as input into Equations 7 through 10.

These indices suggest nontrivial levels of correlation among trait and among method factors. This suggests that there may be additional covariance among trait-related effects and among method-related effects that is not fully accounted for by the person main effect. As noted earlier, whether one considers this unaccounted for covariance as a source of bias (particularly in estimates of person main effect variance) depends on how narrowly one defines the composition of individual variance components being estimated.

Multivariate G-Theory Analyses

To evaluate these data via multivariate G-theory, we utilized the CFA parameterization described in Figure 3. The LISREL syntax used for these analyses is presented in the appendix, and the variance–covariance components estimated by these models are presented in Table 6. This model demonstrated substantially better fit relative to the univariate model ($\chi^2 = 111.927$, $p < .001$; RMSEA = .083; NNFI = .945; CFI = .949; SRMR = .081; GFI = .933); however, this should not be at all surprising as the multivariate model presents a finer grained averaging of the correlation matrix.

There are several noteworthy features regarding the results presented in Table 6. First, the proportion of observed variance in any given trait–method unit that was attributable to the person main effects *for each trait* varied by trait (ranging from .181 for T1 to .237 for T3). Also note that these latent person effects were correlated at a very high level (.892 for T2-T3 to .928 for T1-T3). The latter finding reinforces the results of the univariate G-theory analysis that suggest discrimination among traits was quite poor.

Discussion

To our knowledge, no published work to date provides a parsimonious, jargon-free treatment of G-theory for organizational researchers faced with modeling MTMM data. Our primary goal in the present work was to provide a clear and direct overview of the connections between G-theory, MTMM correlations, traditional notions of convergent and discriminant validity, and CFA. Besides simply making these connections salient for univariate G-theory, we also clarify how multivariate G-theory models relate to these concepts and show their value and limitations for organizational researchers faced with modeling MTMM data. This contribution is particularly important given that treatments of multivariate G-theory have been nonexistent in the organizational research literature yet relevant to the analysis of MTMM data. Given the myriad of problems associated with traditional

Table 6. Multivariate Generalizability Theory Results

		Variance and Covariance Components		
		Trait 1	Trait 2	Trait 3
Σp,pt	Trait 1	.170 (.181)	.911	.928
	Trait 2	.154	.168 (.199)	.892
	Trait 3	.160	.153	.175 (.237)
Σpm,r	Trait 1	.767 (.819)	.642	.469
	Trait 2	.462	.675 (.801)	.517
	Trait 3	.308	.319	.563 (.763)

Note: Σp,pt = variance–covariance matrix of Person Main Effect + Person × Trait interaction variance components (i.e.,$\sigma^2_{p,pti}$ from Table 3) and covariance components (i.e., $\sigma_{p,pti,p,pti'}$ from Table 3); Σpm,r = variance–covariance matrix of Person × Method + Residual Effect variance components (i.e., $\sigma^2_{pm,ri}$ from Table 3) and covariance components (i.e., $\sigma_{pm,ri,pm,ri'}$ from Table 3). In each of the aforementioned matrices, covariance components are listed below the diagonal, correlation components are listed above the diagonal, and variance components are listed on the diagonal. Values listed in parentheses are proportions of observed score variance in trait–method units attributable to the given variance component.

CFA-based approaches to modeling MTMM data, as well as the high regard in which G-theory methods are held in the broader psychometric literature (e.g., Haertel, 2006; Putka & Sackett, 2010), we believe this work can serve as a valuable resource for organizational researchers.

Moreover, utilizing the G-theory methods recommended here as a starting point addresses a key problem that has emerged for traditional CFA-based approaches to modeling MTMM data, namely, issues with convergence and admissibility. The parsimony offered by both the univariate and multivariate G-theory models of MTMM data offers organizational researchers with not only a meaningful starting point grounded in modern psychometric theory but also one that will necessarily converge and produce admissible solutions. Indeed, we have shown how univariate and multivariate G-theory model parameters can be estimated without even invoking the CFA framework. It is for this reason we see the parsimonious nature of the G-theory models as a key advantage.

In addition, the variance-partitioning approaches such as those presented here offer a great deal of flexibility for dealing with the wide variety of measurement designs that underlie MTMM data in practice. We foresee this flexibility being useful in two specific ways: (a) handling designs that involve multiple facets of measurement (e.g., the use of multiple exercises and raters in assessment centers; Arthur et al., 2000; Bowler & Woehr, 2009; Lievens, 2001) and (b) handling designs in which one or more of the measurement factors is *not* crossed with persons (e.g., nested or ill-structured rating designs; Putka, Lance, Le, & McCloy, 2011; Putka et al., 2008). Given the convergence and admissibility problems associated with traditional MTMM CFA models, we posit that extending MTMM CFA models to multitrait-multimethod-multirater (MTMMMR) data may prove even more problematic and the need to start with simpler specifications (such as those offered by G-theory models) to be even more necessary in such contexts. As such, we look forward to future application of G-theory models in such areas.

Conclusion

The ANOVA-based variance decomposition that underlies G-theory was once widely recognized as a potentially fruitful approach to the evaluation of MTMM data (e.g., King et al., 1980; Schmitt & Stults, 1986). Despite the continued popularity of these approaches for the analyses of MTMM and MTMMMR data in educational measurement (e.g., Cronbach et al., 1997; Lane & Stone, 2006; Shavelson et al., 1993; Webb et al., 2000), they have not been widely applied in organizational research. Furthermore, as we have discussed throughout this article, confusion is still apparent in the organizational sciences with regard to how such methods are best implemented and the meaning of the resulting variance components.

The decline of ANOVA-based approaches for analyzing MTMM data in organizational research coincided with the development of CFA-based strategies in the mid-1980s. Over time, CFA emerged as the favored method, as it provided what appeared to be—at least initially—solutions to some of the problems on which ANOVA methods were heavily critiqued (e.g., lack of estimates specific to trait–method units, lack of ready means for testing assumptions). However, over the past few decades, organizational researchers have realized that the use of traditional MTMM CFA models appears to suffer from a number of limitations as well (e.g., Lance et al., 2007). In this article, we showed how univariate and multivariate G-theory models can be framed from the perspective of CFA. This not only provides researchers with a meaningful starting point for examining their data that has clear ties to notions of construct validity but also gives them a clear way for relaxing constraints imposed by the G-theory models as their local situations and theories dictate. We hope that future researchers can take advantage of the connections between G-theory and MTMM data made herein.

Appendix

LISREL Syntax for Fitting the Univariate Generalizability Theory Model to Multitrait–Multimethod Data

```
/*DATA STATEMENT
da ng=1 ni=9 no=347 ma=CM

/*LABELS FOR OBSERVED VARIABLES IN MTMM MATRIX
la
m1t1 m1t2 m1t3 m2t1 m2t2 m2t3 m3t1 m3t2 m3t3

/*MTMM MATRIX
km
1.000
0.675 1.000
0.671 0.705 1.000
0.148 0.125 0.134 1.000
0.234 0.203 0.198 0.608 1.000
0.292 0.203 0.302 0.559 0.581 1.000
0.258 0.190 0.181 0.138 0.134 0.182 1.000
0.220 0.200 0.185 0.148 0.197 0.226 0.771 1.000
0.259 0.226 0.266 0.117 0.123 0.144 0.482 0.518 1.000

/*STANDARD DEVIATIONS  OF OBSERVED VARIABLES
sd
0.905 0.886 0.857 0.904 0.881 0.850 1.084 0.983 0.870
/*MODEL STATEMENT
mo nx=9 nk=7 ph=di,fr lx=fu,fi td=di

/*LABELS FOR LATENT VARIALBES
lk
t1 t2 t3 m1 m2 m3 g
/*STARTING VALUES FOR LAMBDA-X MATRIX
st 1.0 lx (1,1) lx (1,4) lx (1,7)
st 1.0 lx (2,2) lx (2,4) lx (2,7)
```

```
st 1.0 lx (3,3) lx (3,4) lx (3,7)
st 1.0 lx (4,1) lx (4,5) lx (4,7)
st 1.0 lx (5,2) lx (5,5) lx (5,7)
st 1.0 lx (6,3) lx (6,5) lx (6,7)
st 1.0 lx (7,1) lx (7,6) lx (7,7)
st 1.0 lx (8,2) lx (8,6) lx (8,7)
st 1.0 lx (9,3) lx (9,6) lx (9,7)

/*STARTING VALUES FOR PHI MATRIX
st 1.0 ph (1,1)
st 1.0 ph (2,2)
st 1.0 ph (3,3)
st 1.0 ph (4,4)
st 1.0 ph (5,5)
st 1.0 ph (6,6)
st 1.0 ph (7,7)

/*FIXED PARAMETERS FOR LAMBDA-X MATRIX
fi lx (1,1) lx (1,4) lx (1,7)
fi lx (2,2) lx (2,4) lx (2,7)
fi lx (3,3) lx (3,4) lx (3,7)
fi lx (4,1) lx (4,5) lx (4,7)
fi lx (5,2) lx (5,5) lx (5,7)
fi lx (6,3) lx (6,5) lx (6,7)
fi lx (7,1) lx (7,6) lx (7,7)
fi lx (8,2) lx (8,6) lx (8,7)
fi lx (9,3) lx (9,6) lx (9,7)

/*PARAMETERS HELD EQUAL IN THETA-DELTA MATRIX
eq td (1,1) td (2,2) td (3,3) td (4,4) td (5,5) td (6,6) td (7,7) td (8,8) td (9,9)

/*PARAMETERS HELD EQUAL IN PHI MATRIX
eq ph (1,1) ph (2,2) ph (3,3)
eq ph (4,4) ph (5,5) ph (6,6)

/*OUTPUT STATEMENT
ou me=ml rs mi sc ad=off it=5000 nd=3
```

LISREL Syntax for Fitting Multivariate Generalizability Theory Model to Multitrait–Multimethod Data

```
/*DATA STATEMENT
da ng=1 ni=9 no=347 ma=CM

/*LABELS FOR OBSERVED VARIABLES IN MTMM MATRIX
la
m1t1 m1t2 m1t3 m2t1 m2t2 m2t3 m3t1 m3t2 m3t3

/*MTMM MATRIX
```

```
km
1.000
0.675 1.000
0.671 0.705 1.000
0.148 0.125 0.134 1.000
0.234 0.203 0.198 0.608 1.000
0.292 0.203 0.302 0.559 0.581 1.000
0.258 0.190 0.181 0.138 0.134 0.182 1.000
0.220 0.200 0.185 0.148 0.197 0.226 0.771 1.000
0.259 0.226 0.266 0.117 0.123 0.144 0.482 0.518 1.000

/*STANDARD DEVIATIONS  OF OBSERVED VARIABLES
sd
0.905 0.886 0.857 0.904 0.881 0.850 1.084 0.983 0.870

/*MODEL STATEMENT
mo nx=9 nk=3 ph=sy lx=fu,fi td=sy

/*LABELS FOR LATENT VARIALBES
lk
t1 t2 t3

/*PHI PATTERN MATRIX
pa ph
1
1 1
1 1 1

/*THETA-DELTA PATTERN MATRIX
pa td
1
1 1
1 1 1
0 0 0 1
0 0 0 1 1
0 0 0 1 1 1
0 0 0 0 0 0 1
0 0 0 0 0 0 1 1
0 0 0 0 0 0 1 1 1

/*STARTING VALUES FOR LAMBDA-X MATRIX
st 1.0 lx (1,1)
st 1.0 lx (2,2)
st 1.0 lx (3,3)
st 1.0 lx (4,1)
st 1.0 lx (5,2)
st 1.0 lx (6,3)
st 1.0 lx (7,1)
st 1.0 lx (8,2)
st 1.0 lx (9,3)
```

```
/*PARAMETERS CONSTRAINED TO ONE IN LAMBDA-X MATRIX
va 1.0 lx (1,1)
va 1.0 lx (2,2)
va 1.0 lx (3,3)
va 1.0 lx (4,1)
va 1.0 lx (5,2)
va 1.0 lx (6,3)
va 1.0 lx (7,1)
va 1.0 lx (8,2)
va 1.0 lx (9,3)
/*PARAMETERS HELD EQUAL IN THETA-DELTA MATRIX
eq td (1,1) td (4,4) td (7,7)
eq td (2,2) td (5,5) td (8,8)
eq td (3,3) td (6,6) td (9,9)
eq td (2,1) td (5,4) td (8,7)
eq td (3,1) td (6,4) td (9,7)
eq td (3,2) td (6,5) td (9,8)
/*PARAMETERS CONSTRAINED  TO ZERO IN THETA-DELTA MATRIX
va 0.0 td (1,4) td (1,5) td (1,6) td (1,7) td (1,8) td (1,9)
va 0.0 td (2,4) td (2,5) td (2,6) td (2,7) td (2,8) td (2,9)
va 0.0 td (3,4) td (3,5) td (3,6) td (3,7) td (3,8) td (3,9)
va 0.0 td (4,7) td (4,8) td (4,9)
va 0.0 td (5,7) td (5,8) td (5,9)
va 0.0 td (6,7) td (6,8) td (6,9)
/*OUTPUT STATEMENT
ou me=ml rs mi sc ad=off it=5000 nd=3
```

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. The number of traits and methods used here is arbitrary and is simply for illustration. We chose three to draw a clear link to Campbell and Fiske's (1959) seminal work on the analysis of multitrait–multimethod data.
2. The notation "ptm,r" is used to reflect the fact that the three-way Person × Trait × Method interaction is confounded with residual error.
3. It is unnecessary to specify the three-way interaction or residual term in these models. SAS and SPSS will automatically generate an estimate for residual variance, which is completely confounded with variance due to the three-way interaction effect. Further details on the SAS and SPSS procedures for estimating variance components can be found in SAS and SPSS help documentation.
4. SAS's Help and Documentation for the MIXED procedure provides a formula for estimating the memory requirements needed to estimate variance components for a given design.
5. Equations 7 through 10 produce a set of *standardized* variance components, that is, variance components that sum to 1 (i.e., $\sigma^2_{\text{expected observed}} = 1$). Such standardization eases interpretation of variance components

in that each component can be viewed as the proportion of expected observed score variance on a trait–method unit that is accounted for by the given component.

6. Although fitting generalizability theory (G-theory) models within a confirmatory factor analysis framework will produce overall indices of model fit, as we note later researchers will need to exercise caution when considering these indices because they may reflect types of misfit that have little or no bearing on the accuracy of G-theory model parameter estimates.
7. The multivariate G-theory model as applied here does not attempt to tease apart $\sigma^2{}_p$ and $\sigma^2{}_{pt}$. We have subscripted these components with "p,pt" to signify the presence of confounded effects relative to the univariate G-theory model. Similarly, the multivariate G-theory model as applied here does not attempt to tease apart $\sigma^2{}_{pm}$ and $\sigma^2{}_r$; thus, we have subscripted these components with "pm,r" to signify the presence of confounded effects relative to the univariate G-theory model.
8. The covariances among latent residual effects for trait–method units that do not share a method in common are assumed to be zero.

References

Arthur, W., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management*, *26*, 813-835.

Baltes, B. B., Bauer, C. B., Bajdo, L., & Parker, C. P. (2002). The use of multitrait–multimethod data for detecting nonlinear relationships: The case of psychological climate and job satisfaction. *Journal of Business and Psychology*, *17*, 3-17.

Bell, J. F. (1985). Generalizability theory: The software problem. *Journal of Educational Behavior and Statistics*, *10*, 19-29.

Bost, J. E. (1995). The effects of correlated errors on generalizability and dependability coefficients. *Applied Psychological Measurement*, *19*, 191-203.

Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic investigation of assessment center construct validity. *Journal of Applied Psychology*, *91*, 1114-11124.

Bowler, M. C., & Woehr, D. J. (2009). Assessment center construct-related validity: Stepping beyond the MTMM matrix. *Journal of Vocational Behavior*, *75*, 173-182.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems II: Effects of inequality of variance and of correlation between errors in the two way classification. *Annals of Mathematical Statistics*, *29*, 885-891.

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer-Verlag.

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

Cresswell, S. L., & Eklund, R. C. (2006). The convergent and discriminant validity of burnout measures in sport: A multi-trait/multi-method analysis. *Journal of Sport Sciences*, *24*, 209-220.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements. *Theory of generalizability for scores and profiles.* New York: John Wiley.

Cronbach, L. J., Linn, R. L., & Brennan, R. L,, & Haertel, E. H. (1997). Generalizability analyses for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, *57*, 373-399.

DeShon, R. P. (1995, May). Restricted maximum likelihood estimation of variance components in generalizability theory: Overcoming balanced design requirements. Paper presented at the 10th annual conference of the Society of Industrial and Organizational Psychology, Orlando, FL.

DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods*, *3*, 412-423.

DeShon, R. P. (2002). Generalizability theory. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations* (pp. 189-220). San Francisco: Jossey-Bass.

Dickinson, T. L. (1986). Performance ratings: Designs for evaluating their validity and accuracy (AFHRL-TP-86-15). Brooks Air Force Base, TX: Air Force Systems Command.

Guilford, J. P. (1954). *Psychometric methods.* New York: John Wiley.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: American Council on Education and Praeger Publishers.

Halbesleben, J. R. B., & Demerouti, E. (2005). The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work and Stress*, *19*, 208-220.

Hoffman, D. J., & Woehr, D. J. (2009). Disentangling the meaning of multisource feedback source and dimension factors. *Personnel Psychology*, *62*, 735-765.

Jackson, S., & Brashers, D. E. (1994). *Random factors in ANOVA.* Thousand Oaks, CA: Sage.

Kane, M. (1996). The precision of measurements. *Applied Measurement in Education*, *9*, 355-379.

Kane, M. (2002). Inferences about variance components and reliability–generalizability coefficients in the absence of random sampling. *Journal of Educational Measurement*, *39*, 165-181.

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

King, L., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, *65*, 507-516.

Kraiger, K., & Teachout, M. S. (1990). Generalizability theory as construct-related evidence of the validity of performance ratings. *Human Performance*, *3*, 19-35.

Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, *89*, 377-385.

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.

Lance, C. E., Woehr, D. J., & Meade, A. W. (2007). Case study: A Monte Carlo investigation of assessment center construct validity models. *Organizational Research Methods*, *10*, 430-448.

Lane, S., & Stone, C. (2006). Performance assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 387-431). Westport, CT: American Council on Education and Praeger Publishers.

Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior*, *22*, 203-221.

Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, *87*, 675-686.

Marcoulides, G. A. (1990). An alternative method for estimating variance components in generalizability theory. *Psychological Reports*, *66*, 102-109.

Marcoulides, G. A. (1996). Estimating variance components in generalizability theory: The covariance structure analysis approach. *Structural Equation Modeling*, *3*, 290-299.

Marsh, H. W. (1989). Confirmatory factor analysis 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 alternate models. *Applied Psychological Measurement*, *15*, 47-70.

Marsh, H. W., Byrne, B. M., & Craven, R. (1993). Overcoming problems in confirmatory factor analyses of MTMM data: The correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research*, *27*, 489-507.

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.

Mitte, K., & Kämpfe, N. (2008). Personality and the four faces of positive affect: A multitrait–multimethod analysis using self- and peer-report. *Journal of Research in Personality*, *42*, 1370-1375.

Putka, D. J., Lance, C. E., Le, H., & McCloy, R. A. (2011). A cautionary note on modeling multitrait–multirater data arising from ill-structured measurement designs. *Organizational Research Methods, 14,* 503-529.

Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology*, *93*, 959-981.

Putka, D. J., & McCloy, R. A. (2008, February). *Estimating variance components in SPSS and SAS: An annotated reference guide.* Alexandria, VA: Human Resources Research Organization. Retrieved from http://www.humrro.org/corpsite/sites/default/files/dputkaFiles/Estimating Variance Components in SPSS and SAS.pdf

Putka, D. J., & Sackett, P. R. (2010). Reliability and validity. In J.L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 9-49). New York: Routledge.

Raykov, T., & Marcoulides, G. A. (2006). Estimation of generalizability coefficients via a structural equation modeling approach to scale reliability estimation. *International Journal of Testing*, *6*, 81-95.

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

Scullen, S. E. (1999). Using confirmatory factor analysis of correlated uniquenesses to estimate method variance in multitrait–multimethod matrices. *Organizational Research Methods*, *2*, 275-292.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* New York: John Wiley.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, *30*, 215-232.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage.

Smith, P. L., & Luecht, R. M. (1992). Correlated effects in generalizability studies. *Applied Psychological Measurement*, *16*, 229-235.

Stanley, J. C. (1961). Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, *26*, 205-219.

Webb, N. M, Schlackman, J., & Sugrue, B. (2000). The dependability and interchangeability of assessment methods in science. *Applied Measurement in Education*, *13*, 277-301.

Webb, N. M., & Shavelson, R. J. (1981). Multivariate generalizability of general educational developmental ratings. *Journal of Educational Measurement*, *18*, 13-22.

Woehr, D. J., Sheehan, M. K., & Bennett, W., Jr. (2005). Assessing measurement equivalence across rating sources: A multitrait–multimethod approach. *Journal of Applied Psychology*, *90*, 592-600.

Bios

David J. Woehr is a professor in the Department of Management in the Belk College of Business at the University of North Carolina Charlotte. Dr. Woehr received the BA degree in Psychology from Trinity University and the MS and PhD degrees in Industrial and Organizational Psychology from the Georgia Institute of Technology. Dr. Woehr's research interests are in the areas of performance measurement, managerial assessment, and applied psychometrics. His research has appeared in the *Journal of Applied Psychology, Personnel Psychology, Organizational Behavior and Human Decision Processes, Organizational Research Methods, Educational and Psychological Measurement, Journal of Vocational Behavior*, and others.

Dan J. Putka is a Principal Staff Scientist in the Personnel Selection and Development Program at the Human Resources Research Organization (HumRRO). His work involves the development and evaluation of selection systems for clients in the private and public sector, as well as the study and evaluation of personnel issues and policy for the U.S. military. Dan's published research has appeared in book chapters, as well as journals such as *Organizational Research Methods*, the *Journal of Applied Psychology*, and SIOP's *Industrial and Organizational Psychology: Perspectives on Science and Practice*. His research has addressed myriad issues in applied psychometrics, including reliability estimation, modeling of error in ratings, content-oriented validation, differential prediction, and multimodel inference. He received his PhD in industrial and organizational psychology, with a specialization in quantitative methods, from Ohio University.

Mark C. Bowler is an Assistant Professor in the Industrial/Organizational Psychology program at East Carolina University. He earned his doctorate in Industrial/Organizational Psychology from the University of

Tennessee. His primary research interests are in (1) the design and implementation of assessment centers with a focus on the statistical principles relating to the evaluation of their construct-related validity and (2) counterproductive work behaviors. His research in these areas has been published in journals such as *Group and Organization Management, Journal of Applied Psychology,* and *Journal of Vocational Behavior.*