A Multilevel Multitrait-Multimethod Analysis

Cora J. M. Maas, Gerty J. L. M. Lensvelt-Mulders, and Joop J. Hox
Department of Methodology and Statistics, Utrecht University, The Netherlands

Abstract. The classical multitrait-multimethod (MTMM) matrix can be viewed as a two-dimensional cross-classification of traits and methods. Beside commonly used analysis methods such as structural equation modeling and generalizability theory, multilevel analysis offers attractive possibilities. If the focus is only on analyzing classical MTMM data, the multilevel approach has no surplus value, because the resulting model is equivalent to a confirmatory factor model with additional restrictions imposed by the multilevel parameterization. However, if the data contain further complexities, such as additional information on the traits or persons, multilevel analysis of MTMM data offers a flexible analysis tool with more possibilities than the other approaches.

Keywords: multitrait-multimethod matrix, multilevel analysis, structural equation modeling, confirmatory factor analysis

The classical multitrait-multimethod (MTMM) matrix can be viewed as a two-dimensional cross-classification of traits and methods. Several approaches have been proposed to analyze such data, notably structural equation models, generalizability theory, and even facet analysis. Since recent multilevel modeling software can also analyze cross-classified structures, multilevel analysis can be added to this list.

This study investigates the specific advantages and disadvantages of the multilevel approach for analyzing MTMM data. First, in the next section, the MTMM design for the assessment of construct validity is described using a small example data set. The classic Campbell-Fiske criteria, the commonly used confirmatory factor analysis (CFA) and the multilevel analysis for MTMM data, are discussed. For a standard MTMM matrix it is shown that the multilevel approach can be viewed as a confirmatory factor model with additional restrictions.

In the next section a more complicated data set (a part of the German Big 5 MTMM data set, Eid, Lischetzke, Nussbeck, & Geiser, 2004; Eid, Lischetzke, & Trierweiler, 2001) is used to show the advantages of the multilevel approach. The construct validity of measures of three of the Big 5 factors, namely extraversion, neuroticism, and conscientiousness (Ostendorf, 1990), are computed, following a MTMM multilevel approach. The models and results are reported and the attractiveness of the multilevel approach for complicated data structures is discussed. The report ends with conclusions and a discussion, whereby the center of interest lies in the differences between the multilevel approach to MTMM data and the more common CFA approach.

The MTMM Approach to Validity

The MTMM design, originally suggested by Campbell and Fiske (1959), is a design in which several distinct characteristics or traits are measured using several distinct measurement methods. Marsh and Grayson (1995, p. 177) describe traits as “attributes such as multiple abilities, attitudes, behaviors, or personality characteristics”, whereas methods “refer broadly to multiple test forms, methods of assessment, raters, or occasions”. Commonly, each combination of trait and method defines a single measure, which can be a single question or item, or a scale. As an example we use a small data set, about life satisfaction data, presented by Hox (1995). These data are based on an MTMM design with three traits: satisfaction with housing, income, and health, and three methods: questions based on the graphical format of a ladder, faces, and circles. Using a mail survey, data were collected from 498 respondents in The Netherlands (Hox, 1986). This produces an MTMM correlation matrix consisting of $3 \times 3 = 9$ measures.

MTMM designs are used to evaluate construct validity, which requires both high convergent and discriminant validity. Convergent validity implies a high overlap between alternative measures that draw on the same construct (Hoyle, Harris, & Judd, 2002, p. 90). Discriminant validity implies that a measure should not correlate highly with other measures that use the same measurement method but draw on different constructs (Hoyle et al., 2002, p. 92). In addition, convergent validity requires small method effects (Marsh & Grayson, 1995, p. 177).

Campbell-Fiske Criteria

Campbell and Fiske (1959) suggest examining the correlation matrix of the measures defined by different traits and methods. This matrix can be divided in different sections. For good convergent validity, the heteromethod-monotrait correlations should be high, and the monomethod-heterotrait correlations should be low. For good discriminant validity, the correlations between different traits should be reasonably...
low, whether they are measured with the same or a different method.

Table 1 shows the correlations between the nine measures. Since there was a small amount of missing data, these correlations were estimated using the expectation-maximization algorithm in SPSS. The correlations between measures for the same construct based on different methods are given in bold. It is easy to see that the Campbell-Fiske criteria are well met in this MTMM matrix. The highest correlations by far are between heteromethod-monotrait measures. The correlations between traits are in the range denoted *medium effect* by Cohen (1988). The highest correlation between traits is between satisfaction with house and satisfaction with income, which is intuitively understandable.

### CFA

The Campbell-Fiske guidelines provide a diagnostic evaluation of the convergent and discriminant validity of constructs, both of which are an aspect of the construct validity. A more formal approach based on CFA was proposed by Jöreskog (1974). CFA can be used to decompose the underlying factors in the MTMM matrix. Trait factors are defined by different measures of the same trait, and method factors are defined by different constructs assessed with the same method. The basic assumptions are that each measure loads on only one trait and one method factor, and that the covariances between trait and method factors are zero. Large factor loadings for the traits indicate high convergent validity, large method factor loadings imply measurement bias due to method effects, and high correlations between trait factors indicate a lack of discriminant validity. Correlations between method factors are optional. Figure 1 shows the basic MTMM model for the satisfaction data.

As Figure 1 shows, the scale of the trait and method factors is set by constraining their variance to one. The model has an excellent fit ($\chi^2(498) = 19.5$, $df = 15$, $p = .19$; $GFI = 0.99$, $CFI = 1.00$, and $RMSEA = 0.03$). All parameters in the model are significant, except some of the covariances between the method factors. Table 2 shows the standardized coefficients for the CFA.

### Table 1. Correlations, means, and standard deviations satisfaction data

<table>
<thead>
<tr>
<th>Items</th>
<th>L-H</th>
<th>L-I</th>
<th>L-H</th>
<th>F-H</th>
<th>H-I</th>
<th>F-H</th>
<th>C-H</th>
<th>C-I</th>
<th>C-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ladder-house</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ladder-income</td>
<td>.359</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ladder-health</td>
<td>.240</td>
<td>.299</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Faces-house</td>
<td>.837</td>
<td>.284</td>
<td>.179</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Faces-income</td>
<td>.254</td>
<td>.862</td>
<td>.249</td>
<td>.298</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Faces-health</td>
<td>.196</td>
<td>.208</td>
<td>.897</td>
<td>.223</td>
<td>.252</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Circles-house</td>
<td>.832</td>
<td>.330</td>
<td>.190</td>
<td>.851</td>
<td>.293</td>
<td>.197</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Circles-income</td>
<td>.247</td>
<td>.833</td>
<td>.250</td>
<td>.272</td>
<td>.864</td>
<td>.225</td>
<td>.361</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Circles-health</td>
<td>.169</td>
<td>.204</td>
<td>.848</td>
<td>.173</td>
<td>.218</td>
<td>.864</td>
<td>.208</td>
<td>.254</td>
<td>1</td>
</tr>
<tr>
<td>Means</td>
<td>5.30</td>
<td>4.87</td>
<td>5.35</td>
<td>5.43</td>
<td>4.96</td>
<td>5.40</td>
<td>4.54</td>
<td>4.17</td>
<td>4.59</td>
</tr>
<tr>
<td>Standard deviations</td>
<td>1.305</td>
<td>1.338</td>
<td>1.439</td>
<td>1.209</td>
<td>1.278</td>
<td>1.377</td>
<td>1.124</td>
<td>1.231</td>
<td>1.226</td>
</tr>
</tbody>
</table>

As can be derived from Table 2 the factor loadings are much higher for the latent traits than for the measurement methods. This means that the trait effect on the items is much larger than the method effect, and large enough to indicate high convergent validity. The loadings for the methods are of moderate size, indicating that there are small method effects. The correlations between the trait factors are moderate (house-income = .33, health-income = .26, health-house = .22), meaning that the traits have acceptable discriminant validity.

### The Multilevel Approach

The path diagram in Figure 1 can be described by the following CFA equation:

$$Y_{ij} = I_i + t_{1i} \theta_{ij} + t_{2i} \theta_{k} + t_{3i} \theta_{j} + m_{1i} \mu_{ij} + m_{2i} \mu_{k} + m_{3i} \mu_{j} + e_{ij},$$

(1)

where $Y_{ij}$ is the score of subject $j$ on item $i$, $I_i$ is the intercept of item $i$, $t_{1i}$ is the loading of item $i$ on trait $k$, $\theta_{kj}$ is the score of subject $j$ on trait $k$, $m_{1i}$ is the loading of item $i$ on method $k$, $\mu_{kj}$ is the score of subject $j$ on method $k$, and $e_{ij}$ is the residual error of subject $j$ for item $i$. The factor loading of indicators who do not belong to a specific trait or method are constrained to zero. The intercepts $I_i$ are implicit in Figure 1, and they are necessary because the incomplete data are analyzed using Full Information Maximum Likelihood (FIML; Arbuckle, 1996). The scale of the factors is defined by constraining their means to zero and the variances to one, and all intercepts and all unconstrained loadings are freely estimated.

To use multilevel analysis on these MTMM data, the data file must be restructured. In the CFA model, the data are treated as a multivariate response vector. Multilevel regression analysis is in principle a univariate model. Multivariate data can be analyzed in the multilevel regression framework by adding an additional level for the variables. So, the nine MTMM variables are turned into nine cases nested within the subjects. If a specific response variable is missing for a specific case, only that one observation is deleted at the variables level, and the other variables for that
subject are retained. Thus, all available data for each subject are used in the Maximum Likelihood estimation, a procedure equivalent to the FIML treatment of incomplete data in CFA.

To analyze the MTMM design in the restructured data file, three trait dummies and three method dummies are added to the data. A standard multilevel analysis of the variables within subjects data, using the dummies to indicate to which trait and method factor each item belongs, is described by the following equation, using a notation similar to Equation 1:

$$Y_{ij} = I_j + \theta_1 T_{1i} + \theta_2 T_{2i} + \theta_3 T_{3i} + \mu_1 M_{1i} + \mu_2 M_{2i} + \mu_3 M_{3i} + \epsilon_{ij}.$$  

Equation 2 makes clear that the standard multilevel model treats the variation in item means by estimating one person-level intercept that varies across subjects and including an item level error term $\epsilon_{ij}$ to allow for differences in item means. To make the multilevel model more similar to the structural equation model, the intercept is removed from the model, and nine dummy variables $D_1 \ldots D_9$ are added to indicate each of the nine MTMM items. The regression coefficients $d_i$ for the nine dummy variables are allowed to vary across items. This model is given by Equations 3 and 4:

$$Y_{ij} = \theta_1 T_{1i} + \theta_2 T_{2i} + \theta_3 T_{3i} + \mu_1 M_{1i} + \mu_2 M_{2i} + \mu_3 M_{3i} + I_i D_i + u_{ij}. \quad (3)$$

$$\beta_i = I_i + u_{ij}. \quad (4)$$

Combined into one equation:

$$Y_{ij} = \theta_1 T_{1i} + \theta_2 T_{2i} + \theta_3 T_{3i} + \mu_1 M_{1i} + \mu_2 M_{2i} + \mu_3 M_{3i} + I_i D_i + u_{ij} D_i. \quad (5)$$

For clarity in Equations 3 and 5 the equation is written explicitly for the traits and methods, while the dummy part is written in vector notation for brevity. Compared to

Table 2. Results of CFA on satisfaction data

<table>
<thead>
<tr>
<th>Items</th>
<th>Trait loading*</th>
<th>Method loading*</th>
<th>Trait variance</th>
<th>Method variance</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>House-ladder</td>
<td>.902</td>
<td>.272</td>
<td>.814</td>
<td>.074</td>
<td>.814</td>
<td>.888</td>
</tr>
<tr>
<td>House-faces</td>
<td>.927</td>
<td>.318</td>
<td>.859</td>
<td>.101</td>
<td>.859</td>
<td>.960</td>
</tr>
<tr>
<td>House-circles</td>
<td>.917</td>
<td>.210</td>
<td>.841</td>
<td>.044</td>
<td>.841</td>
<td>.885</td>
</tr>
<tr>
<td>Income-ladder</td>
<td>.910</td>
<td>.181</td>
<td>.828</td>
<td>.033</td>
<td>.828</td>
<td>.861</td>
</tr>
<tr>
<td>Income-faces</td>
<td>.947</td>
<td>.164</td>
<td>.897</td>
<td>.027</td>
<td>.897</td>
<td>.924</td>
</tr>
<tr>
<td>Income-circles</td>
<td>.941</td>
<td>.167</td>
<td>.830</td>
<td>.028</td>
<td>.830</td>
<td>.858</td>
</tr>
<tr>
<td>Health-ladder</td>
<td>.940</td>
<td>.249</td>
<td>.884</td>
<td>.062</td>
<td>.884</td>
<td>.946</td>
</tr>
<tr>
<td>Health-faces</td>
<td>.956</td>
<td>.131</td>
<td>.914</td>
<td>.017</td>
<td>.914</td>
<td>.931</td>
</tr>
<tr>
<td>Health-circles</td>
<td>.903</td>
<td>.269</td>
<td>.815</td>
<td>.072</td>
<td>.815</td>
<td>.887</td>
</tr>
</tbody>
</table>

*Standardized regression loadings.
Equation 2, in Equation 5 the item variance is completely modeled by the nine item intercepts that model the mean structure. The dummy variables $D_1 \ldots D_9$ are just a device to refer in the multilevel structure to each of the nine variables. Hence, the nine intercepts $I_i$ for the dummy variables $D_i$ are the same as the intercepts $I_i$ estimated in the CFA, and the nine variances for the residual errors $u_i$ are the same as the residual measurement errors $e_i$ in Equation 1. In the multilevel model, all covariances between the intercepts $I_i$ are constrained to zero, covariances between the regression coefficients for traits on the one side and methods on the other side are also constrained to zero. When Equations 1 and 3 are compared, it is clear that the regression coefficients $\theta$ and $\mu$ in Equation 3, which vary across persons, are the equivalent of the latent variables $\theta$ and $\mu$ in Equation 1, and the dummy variables $T$ and $M$ in Equation 3 are the equivalent of the factor loadings $t$ and $m$ in Equation 1. There is one important difference. In the CFA the factor loadings $t$ and $m$ are estimated differently for each item, whereas in the multilevel analysis the dummy variables $T$ and $M$ are zero-one which leads to factor loadings that are all equal to one.

The results of the multilevel analysis are given in Table 3.

The multilevel standardized loadings are approximately equal to the mean of the CFA standardized loadings. The same holds for the variance components, the validity and the reliability.

Since the multilevel analysis of these MTMM data is simply a reparameterization of the CFA model, with added constraints of equal factor loadings for each factor, we can obtain the results in Table 3 also using CFA, by adding these constraints. The single factor loading for each factor is then estimated as the square root of the variance of the corresponding regression coefficient in the multilevel analysis. When we specified the multilevel model as a constrained CFA, we obtained the same estimates within two decimal places. Since these constraints define a nested model, we can test whether they are tenable using the chi-square difference test ($\Delta \chi^2$). The restricted model also showed a good fit to the data ($\chi^2(498) = 47.0$, $df = 27$, $p = .01$; GFI = 0.98, CFI = 0.99, and RMSEA = 0.04). However, the $\Delta \chi^2$ test is significant so the unrestricted model should be preferred above the restricted model ($\Delta \chi^2 = 27.5$, $df = 12$, $p < .01$).

### Multilevel Analysis of a Complex Data Set

#### Description of the Data

For the complex data set, a part of the German Big 5 MTMM data set is used (Eid et al., 2001, 2004). This data set does not only contain information about traits and methods, but also additional information about the raters. The three traits of the data set are taken from the Big 5 personality traits: extraversion, neuroticism, and conscientiousness. Each trait is measured by four items with five response categories. There appear to be three methods: self-report (SR) and two peer reports from peer A and peer B (PRA and PRB). Each target person was required to bring two peers to the measurement occasion. In total, data were collected from 482 triplets. To characterize the triplets, several additional questions were asked about their mutual relationship (see Figure 2).

#### Attractiveness of Multilevel Analysis for Complex Data

This data set contains several complexities that are absent from the simple MTMM discussed above. Firstly, in the standard MTMM design we usually have only one measure per trait-method combination, here we have four items in each cell of the MTMM matrix. Secondly, although there are two peers, they are exchangeable and both belong to the same method facet peers. Thirdly, there is background information that we wish to include in the MTMM analysis. In the multilevel context, these complexities can be handled in a straightforward manner. There is no requirement that there is only one measurement per cell, additional measures are simply included at the variables level. Similarly, there is no requirement that each of the two peers is assigned to a separate method facet, they are simply both included in the single peers facet. Finally, when a data set contains additional information, this information can be added to the multilevel regression equation as covariates. For instance, in the German Big 5 data, also the gender of the target person is known. Adding this covariate to the equation gives:

---

**Table 3. Results of multilevel analysis on satisfaction data**

<table>
<thead>
<tr>
<th>Items</th>
<th>Trait loading</th>
<th>Method loading</th>
<th>Trait variance</th>
<th>Method variance</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>House-ladder</td>
<td>.915</td>
<td>.258</td>
<td>.837</td>
<td>.067</td>
<td>.837</td>
<td>.904</td>
</tr>
<tr>
<td>House-faces</td>
<td>.915</td>
<td>.174</td>
<td>.837</td>
<td>.030</td>
<td>.837</td>
<td>.867</td>
</tr>
<tr>
<td>House-circles</td>
<td>.915</td>
<td>.210</td>
<td>.837</td>
<td>.044</td>
<td>.837</td>
<td>.881</td>
</tr>
<tr>
<td>Income-ladder</td>
<td>.923</td>
<td>.258</td>
<td>.852</td>
<td>.067</td>
<td>.852</td>
<td>.919</td>
</tr>
<tr>
<td>Income-faces</td>
<td>.923</td>
<td>.174</td>
<td>.852</td>
<td>.030</td>
<td>.852</td>
<td>.882</td>
</tr>
<tr>
<td>Income-circles</td>
<td>.923</td>
<td>.210</td>
<td>.852</td>
<td>.044</td>
<td>.852</td>
<td>.896</td>
</tr>
<tr>
<td>Health-ladder</td>
<td>.937</td>
<td>.258</td>
<td>.878</td>
<td>.067</td>
<td>.878</td>
<td>.945</td>
</tr>
<tr>
<td>Health-faces</td>
<td>.937</td>
<td>.174</td>
<td>.878</td>
<td>.030</td>
<td>.878</td>
<td>.908</td>
</tr>
<tr>
<td>Health-circles</td>
<td>.937</td>
<td>.210</td>
<td>.878</td>
<td>.044</td>
<td>.878</td>
<td>.922</td>
</tr>
</tbody>
</table>

*Standardized regression loadings.

© 2009 Hogrefe & Huber Publishers
Adding a covariate to the regression equation gives more insight into the explanation of the scores on the three different traits. Like in normal regression analysis, the estimated parameters (i.e., regression coefficients and variance components) are different because they are now conditional on the added covariate.

First, Table 4 gives the standardized regression coefficients for the restricted model, without covariates. As can be derived from Table 4 the factor loadings are much higher for the latent traits than for the measurement methods. This means that the trait effect on the item measures is much larger than the method effect.

Adding the covariate to the model gives a significant effect. Female score higher on the three traits than male (b = 0.25, p < .001). The trait and method loadings change only a minimal amount (maximum difference is 0.02).

Results Multilevel Analysis

Before applying the ML model to the data we first did a CFA in AMOS on a restricted and unrestricted model without covariates. Both models fitted well. The unrestricted model had a $\chi^2$ of 19.70, df = 15, $p = .08$; GFI = 0.78, CFI = 0.99, and RMSEA = 0.026. The restricted model, all factor loadings restricted to 1, had a $\chi^2$ of 47.694, df = 27, $p = .01$; GFI = 0.94, CFI = 0.97, and RMSEA = 0.04. Since the $\Delta \chi^2$ test is significant ($\Delta \chi^2 = 27.99$, df = 12, $p < .05$), the unrestricted model should be preferred. But as we saw above, this is not the only measure, may be the advantages of ML could outweigh the preference for the unrestricted model.

Conclusion and Discussion

In this paper the confirmatory factor model and the multilevel regression model are compared analyzing MTMM data. It is shown that for a standard MTMM matrix the multilevel approach can be viewed as a confirmatory factor model with additional restrictions on the factor loadings. In the confirmatory factor model separate factor loadings for traits and methods can be estimated, while in the multilevel model equal effects of each trait and method on the outcome variables are assumed. Therefore, if the focus is on analyzing classical MTMM data, the confirmatory factor
model offers the most opportunities. In both data sets analyzed in this paper, the fit of the unrestricted factor model is better than the fit of the restricted factor model, indicating that it should be preferred above the necessarily restricted multilevel model.

However, if the data contain further complexities, such as multiple raters or additional information on the traits or persons, as in the Big 5 example, multilevel analysis of MTMM data offers a flexible analysis tool. The additional information can be added in a straightforward manner to the model, so that not only the construct validity of the traits can be established but also the scores on the different traits can be explained.

It should be noted that the limitations in the multilevel model are best understood as limitations that are the consequence of staying within the framework of multilevel regression analysis, which is intrinsically a univariate method. The multivariate structure is modeled by including an extra variables level, and this particular parameterization results in the limitations discussed. As Mehta and Neale (2005) show, multilevel models can be formulated in the structural equation framework. Structural equation modeling (SEM) is a multivariate model, and multilevel SEM has the potential to combine the strengths of both approaches discussed here. However, current multilevel SEM software is mostly limited to two-level models, which limits the applicability.

References


Cora J. M. Maas

Department of Methodology and Statistics
Faculty of Social Sciences
Utrecht University
P.O. Box 80140
3508, TC Utrecht
The Netherlands
Tel. +31 30 2534594
Fax +31 30 2535797
E-mail c.maas@uu.nl

© 2009 Hogrefe & Huber Publishers

Methodology 2009; Vol. 5(3):72–77