

# Multilevel Multitrait Multimethod Model. Application to the Measurement of Egocentered Social Networks

Lluís Coromina<sup>1</sup>, Germà Coenders<sup>2</sup>, and Tina Kogovšek<sup>3</sup>

## Abstract

Our goal in this paper is to assess reliability and validity of egocentered network data using multilevel analysis (Muthén, 1989, Hox, 1993) under the multitrait-multimethod approach. The confirmatory factor analysis model for multitrait-multimethod data (Werts & Linn, 1970; Andrews, 1984) is used for our analyses.

In this study we reanalyse a part of data of another study (Kogovšek et al., 2002) done on a representative sample of the inhabitants of Ljubljana. The traits used in our article are the name interpreters. We consider egocentered network data as hierarchical; therefore a multilevel analysis is required. We use Muthén's partial maximum likelihood approach, called pseudobalanced solution (Muthén, 1989, 1990, 1994) which produces estimations close to maximum likelihood for large ego sample sizes (Hox & Mass, 2001).

Several analyses will be done in order to compare this multilevel analysis to classic methods of analysis such as the ones made in Kogovšek et al. (2002), who analysed the data only at group (ego) level considering averages of all alters within the ego.

We show that some of the results obtained by classic methods are biased and that multilevel analysis provides more detailed information that much enriches the interpretation of reliability and validity of hierarchical data. Within and between-ego reliabilities and validities and other related quality measures are defined, computed and interpreted.

---

1 Department of Economics, University of Girona. FCEE, Campus Montilivi, 17071 Girona, Spain; lluis.coromina@udg.es

2 Department of Economics, University of Girona. FCEE, Campus Montilivi, 17071 Girona, Spain; germa.coenders@udg.es

3 Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia; tina.kogovsek@guest.arnes.si

## 1 Introduction

Our aim in this article is to assess the quality of measurement in social network analysis. More precisely, we are going to use multilevel factor analysis (Muthén, 1989; Hox, 1993) to assess reliability and validity in egocentered networks. Egocentered networks (also called personal networks) consist of a single individual (usually called ego) with one or more relations defined between him/her and a number of other individuals—the members of his/her personal network—called alters (e.g., Kogovšek et al., 2002). Another common type of network is the complete network, which consists of a group of individuals with one or more relations defined among them.

Usually, several characteristics (variables) are measured which describe the ego's relationships (frequently called ties) with his/her alters and the characteristics of alters themselves. Tie characteristics may involve for instance, the type of relation between the ego and the alter (e.g., partner, boss, co-worker), feelings of closeness or importance, duration of the tie and so on. These kinds of questions are frequently called name interpreters (e.g., Kogovšek et al. 2002).

In this paper we want to estimate the reliability and validity of some frequently used name interpreters. Since the data about the characteristics of ties are used as important explanatory variables in social support research and are, moreover, usually reported only by the ego, it is very important to know to what extent these data are reliable and valid.

With this purpose we will use the Multitrait-Multimethod (MTMM) approach. Several other approaches exist to estimate the quality of a measurement instrument (Saris, 1990a) like the quasi-simplex approach (Heise, 1969) and the repeated multimethod approach (Saris, 1995) but will not be dealt with in this paper.

Many different MTMM models have been suggested in the literature. Among them are the correlated uniqueness model (Marsh, 1989; Marsh & Bailey, 1991); the confirmatory factor analysis (CFA) model for MTMM data (Althausen et al., 1971; Alwin, 1974; Werts & Linn, 1970; Andrews, 1984); the direct product model (Browne, 1984, 1985); and the true score (TS) model for MTMM data (Saris & Andrews, 1991). The MTMM model has rarely been used for measurement quality assessment in social network analysis. Hlebec (1999), Ferligoj and Hlebec (1999) and Kogovšek et al. (2002) used the TS model on network data, in the context of complete networks the first two and of egocentered networks the last. The CFA specification is used in this study, not the TS. However, both models are equivalent (e.g., Coenders & Saris, 2000).

In this study we reanalyse a part of data of another study (Kogovšek et al., 2002) done on a representative sample of the inhabitants of Ljubljana. The traits used in our article are the name interpreters frequency of contact, feeling of

closeness, feeling of importance and frequency of the alter upsetting the ego. The methods used are face to face and telephone interviewing.

We consider egocentered network data as hierarchical, therefore a multilevel analysis is required. Multilevel analysis decomposes the total observed scores at the individual level into a between group (ego) component and a within group component. The sample covariance matrix is also decomposed. With this purpose we use Muthén's approach (Muthén, 1989, 1990, 1994). In the balanced case (each ego has the same number of alters), Muthén's approach provides maximum likelihood (ML) estimates of population parameters (Hox, 1993). In the more common unbalanced case, two estimators exist, the Full Information Maximum Likelihood (FIML), and the Partial Maximum Likelihood approach (MUML), called pseudobalanced solution, too. MUML produces estimations close to ML for large ego sample sizes (Hox and Mass, 2001), which is our case, and so it is the method we use.

The structure of this article is as follows. First we will present the standard CFA MTMM model and interpret the reliability and validity estimates provided. Then we will present the data used and argue for their hierarchical nature. Then the CFA MTMM model will be reformulated as a multilevel model and estimation, testing and interpretation issues will be discussed. Finally, several analyses will be performed in order to compare this multilevel analysis to the classic approaches suggested by Härnqvist (1978) and done in Kogovšek et al. (2002), who analysed the data only at group (ego) level considering averages of all alters within the ego. It will be shown that some of the results obtained by classic methods are biased. Besides, the multilevel approach provides much more detailed information and thus a much richer view on measurement quality.

## **2 Reliability and validity assessment**

### **2.1 Reliability and validity defined**

Reliability can be defined as the extent to which any questionnaire, test or measure produces the same results on repeated experiments. However, a random error will always exist. The repeated measures, will not be exactly the same, but will be consistent to a certain degree. The more consistent the results given by repeated measurements, the higher reliability of the measurement procedure.

In order to have a good quality of measurement, reliability is not enough, but we need validity too. Validity is defined as the extent to which any measure measures what is intended to measure (Carmines & Zeller, 1979:12). Validity is affected by the error called systematic error, which is not random but has a systematic biasing effect on the measurement instruments.

Within construct validity we consider nomological, convergent and discriminant validity. Nomological validity implies that the relationships between measures of different concepts must be consistent with theoretically derived hypotheses concerning the concepts that are being measured (Carmines & Zeller, 1979). Convergent validity refers to common trait variance and is inferred from large and statistically significant correlations between different measures of the same trait. Discriminant validity refers to the distinctiveness of the different traits; it is inferred when correlations among different traits are less than one.

The amount of both random and systematic error present in a measurement can depend on any characteristic of the design of the study, such as data collection mode, questionnaire wording, response scale, type of training of the interviewer, all of which can be broadly considered as methods (Groves, 1989).

## 2.2 MTMM Model

In this study the main concerns are convergent and discriminant validity and reliability. Convergent and discriminant validity of different methods was first assessed in a systematic way by the design that we are going to use, the MTMM design (Campbell and Fiske, 1959). In this design three or more traits (variables of interest) are each measured with three or more methods.

Reliability assessment is based on the classical test theory (Lord & Novick, 1968) whose main equation is:

$$Y_{ij} = S_{ij} + e_{ij} \quad (2.1)$$

where:

- $S$  is the part of the response that would be stable across identical independent repetitions of the measurement process and is called true score (Sarıs & Andrews, 1991).
- $Y_{ij}$  is the response or measured variable  $i$  measured by method  $j$ .
- $e_{ij}$  is the random error, which is related to lack of reliability.

In coherence with the MTMM approach, the stable part is assumed to be the combined result of trait and method:

$$S_{ij} = m_{ij} M_j + t_{ij} T_i \quad (2.2)$$

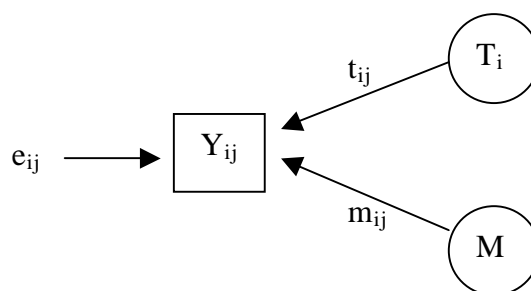
where:

- $T_i$  is the unobserved variable of interest (trait). Related to validity.
- $M_j$  is the variation in scores due to the method. Related to invalidity.
- $m_{ij}$  and  $t_{ij}$  are factor loadings on the method and trait factors respectively

Equations 2.1 and 2.22 constitute the specification of the true score (TS) MTMM model of Saris and Andrews (1991). By substitution we obtain Equation 2.3 which corresponds to the Confirmatory Factor Analysis (CFA) specification of the MTMM model (Andrews, 1984):

$$Y_{ij} = m_{ij} M_j + t_{ij} T_i + e_{ij} \quad (2.3)$$

It can be shown that both models are equivalent (Coenders & Saris, 2000). Equation 2.3 is depicted in Figure 1.



**Figure 1:** Path diagram for the MTMM model for trait ( $T_i$ ) and method ( $M_j$ ).

In this model it is necessary to make some assumptions (Andrews, 1984):

$$\begin{aligned} cov(T_i, e_{ij}) &= 0 \quad \forall_{ij} \\ cov(M_j, e_{ij}) &= 0 \quad \forall_{ij} \\ cov(M_j, T_i) &= 0 \quad \forall_{ij} \\ E(e_{ij}) &= 0 \quad \forall_{ij} \\ cov(M_j, M_{j'}) &= 0 \quad \forall_{j \neq j'} \\ m_{ij} &= 1 \quad \forall_{ij} \end{aligned} \quad (2.4)$$

which imply:

- There is no correlation between the errors and the latent variables, both traits and methods.
- There is no correlation between the traits and the methods. These first two assumptions make it possible to decompose the variance of  $Y_{ij}$  into trait variance  $t_{ij}^2 \text{var}(T_i)$ , method variance  $m_{ij}^2 \text{var}(M_j)$  and random error variance  $\text{var}(e_{ij})$  to assess measurement quality (Schmitt & Stults, 1986).
- The expectation of the random error is zero.
- There is no correlation between methods.
- Method effects are equal within methods. The last two assumptions are not always made. They were suggested by Andrews (1984) and Sherpenzeel (1995) as a means to improve the stability of the model, that is increase the

rate of convergence of the estimation procedures, reduce the rate of appearance of inadmissible solutions (e.g., negative variances) and reduce standard errors (Rindskopf, 1984). Problems in these respects had been often reported in much previous research using the CFA model (Bagozzi & Yi, 1991; Brannick & Spector, 1990; Kenny & Kashy, 1992; Marsh & Bailey, 1991; Saris, 1990b).

Usually at least three methods are required. In this article only two will be used. If only two methods are used, the model with all constraints in Equation 2.4 is still identified but rather unstable and standard errors can get very large. In order to increase the stability of the model, the additional constraint that  $t_{ij}$  are constant within method is considered:

$$t_{ij} = t_{i'j} \quad \forall_{i \neq i'} \quad (2.5)$$

This constraint is reasonable if the relationship between the units of measurement of Method 1 and the units of measurement of Method 2 are constant across traits. This assumption is reasonable if the response scales do not vary across methods (this will be the case in our article) or they vary in the same way for all traits. If we impose this assumption standard errors get much lower; with our data they got 29,7% lower on average.

The definitions of reliability and validity from classical test theory used in Saris and Andrews (1991) for the TS model can also be implemented in the CFA formulation of the model as follows. Reliability is the proportion of variance in  $Y_{ij}$  that is stable across repeated identical measures:

$$Reliability = \frac{Var(S_{ij})}{Var(Y_{ij})} = \frac{m_{ij}^2 Var(M_j) + t_{ij}^2 Var(T_i)}{Var(Y_{ij})} \quad (2.6)$$

and the reliability coefficient is the square root of reliability. Thus, reliability increases not only with true or trait variance, but also with method variance, which also belongs to the stable or repeatable part of the measurements.

Validity, assuming that method is the only source of invalidity, is:

$$Validity = \frac{t_{ij}^2 Var(T_i)}{Var(S_{ij})} = \frac{t_{ij}^2 Var(T_i)}{m_{ij}^2 Var(M_j) + t_{ij}^2 Var(T_i)} \quad (2.7)$$

and the validity coefficient is the square root of validity. Validity is thus the percentage of variance of the true score explained by the trait. As explained before, the true score is the trait effect plus the method effect. Then, we can assess invalidity as 1 minus validity.

Another definition of validity uses the total variance in the denominator of Equation 2.7, thus making reliability be the upper bound of validity. The advantage of the definition used in Saris and Andrews (1991) and presented here is that it makes the range of validity independent of the value of reliability, as validity can be equal to 1 even for unreliable measures. The advantage of the other definition is that validity is understood as an overall measurement quality indicator, as it assesses the percentage of trait variance contained in the total variance of each measure.

1. How frequently are you in contact with this person (personally, by mail, telephone or Internet)?

1 Less than once a year.  
 2 Several times a year.  
 3 About once a month.  
 4 Several times a month.  
 5 Several times a week.  
 6 Every day.

2. How close do you feel to this person? Please describe how close you feel on a scale from 1 to 5, where 1 means not close and 5 means very close.

1	2	3	4	5
<b>Not Close</b>		<b>Very Close</b>		

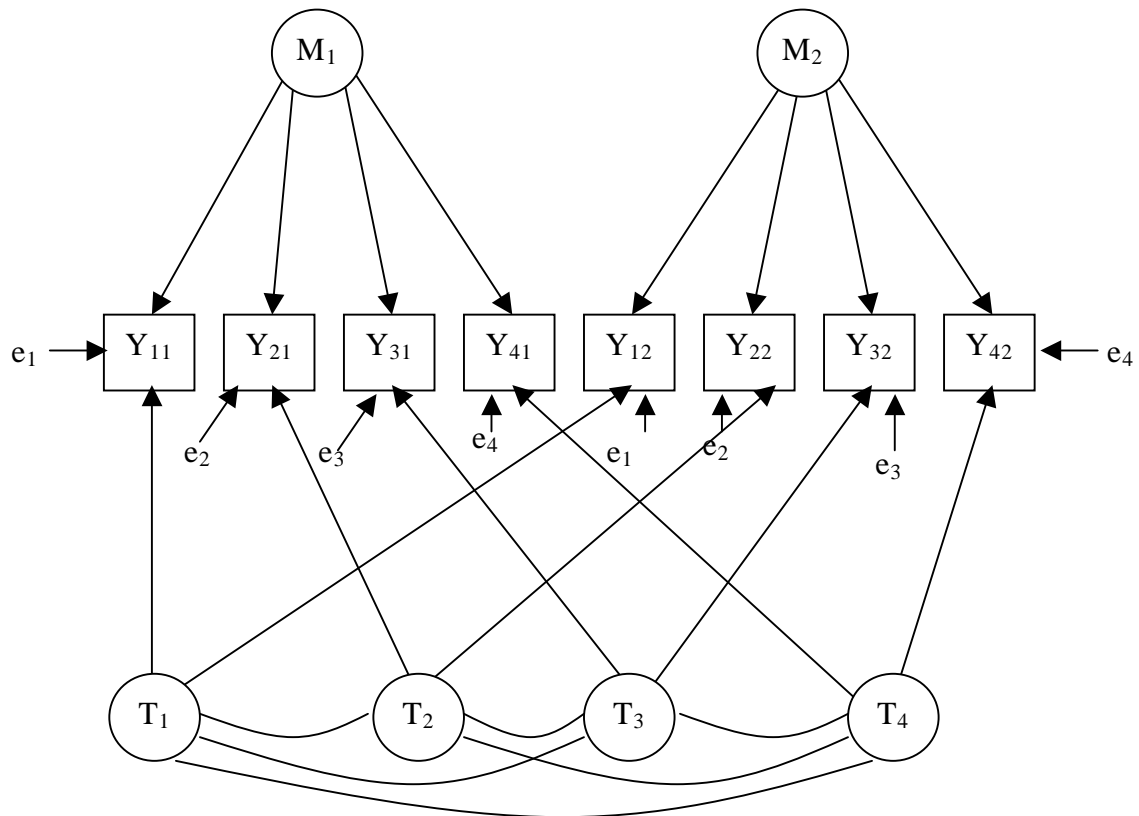
3. How important is this person in your life? Please describe how close you feel on a scale from 1 to 5, where 1 means not important and 5 means very important.

1	2	3	4	5
<b>Not important</b>		<b>Very important</b>		

4. How often does this person upset you?

1 Never.  
 2 Rarely.  
 3 Sometimes.  
 4 Often.

**Figure 2:** Questionnaire.



**Figure 3:** Path diagram of a CFA MTMM model for two methods and four traits.

### 3 Data

The kind of network that we are going to study is known as egocentered network or personal network. It consists of a single individual (usually called ego) with one or more relations defined between him/her and a number of other individuals—the members of his/her personal network, called alters (Kogovšek et al., 2002:2).

First of all, it is necessary to find the ego's network. Name generators are questions for eliciting the names of the ego's network members (alters). Secondly, other questions are used to describe these relationships, such as frequency of contact with the alter, feeling of closeness to the alter, feeling of importance of the relationship and so on. These kinds of questions are frequently called name interpreters. Our aim is to estimate the reliability and validity of some of the very frequently used name interpreters (traits) using different methods.

With this purpose we reanalyse a part of the data of another study (Kogovšek et al., 2002) done on a representative sample of the inhabitants of Ljubljana. The complete study of Kogovšek et al involved several subsamples with several missing data patterns planned by design. In this article we use only one group without missing data, as the aim of the current paper is a different one. The part of



the sample used in this paper consists of  $G=314$  egos who evaluated  $N=1371$  alters. The subset of variables used by us is described below:

*Traits*

$T_1$	Frequency of contact
$T_2$	Feeling of closeness
$T_3$	Feeling of importance
$T_4$	Frequency of the alter upsetting to ego

*Methods*

$M_1$	Face-to-face interviewing
$M_2$	Telephone interviewing

The wording of the name interpreters used in this study is displayed in Figure 2. Figure 3 displays a CFA model for two methods and four traits.

## 4 Multilevel analysis

### 4.1 Model and estimation

Egocentered network data can be considered as hierarchical. An ego chooses the alters according to the name generator questions and therefore the alters are a “part of” the egos. In the hierarchical structure there are the egos at the top of the hierarchy, and all their alters at the bottom. Thus, alters are nested into the egos in what constitutes a nested data structure. Responses to the name interpreter questions constitute the data.

The technique used will be two-level factor analysis, or, more particularly, two-level MTMM models. The lowest level is known as individual level and the highest level is known as group level. Thus, in our case groups will be egos and individuals will be alters.

The mean centred individual scores for group  $g$  and individual  $k$   $Y_{T_{gk}} = Y_{gk} - \bar{Y}$  can be decomposed into a between group component  $Y_{B_g} = \bar{Y}_g - \bar{Y}$  and a within group component  $Y_{W_{gk}} = Y_{gk} - \bar{Y}_g$ . Since both components are independent, the cross product matrices can be aggregated as:

$$\begin{aligned} \sum^G \sum^n (Y_{gk} - \bar{Y})(Y_{gk} - \bar{Y})' = \\ n \sum^G (\bar{Y}_g - \bar{Y})(\bar{Y}_g - \bar{Y})' + \sum^G \sum^n (Y_{gk} - \bar{Y}_g)(Y_{gk} - \bar{Y}_g)' \end{aligned} \quad (4.1)$$

where:

- $\bar{Y}$  is the total average over all alters and egos.
- $\bar{Y}_g$  is average of all alters of the gth ego.
- $Y_{gk}$  is the score on the name interpreter of the kth alter chosen by the gth ego.
- $G$  is the total number of egos.
- $n$  is the number of alters within each ego, assumed to be constant.
- $N=nG$  is the total number of alters.

The sample covariance matrices are obtained when dividing the components in Equation 4.1 by their degrees of freedom:

$$S_W = \frac{\sum^G \sum^n (Y_{gk} - \bar{Y}_g)(Y_{gk} - \bar{Y}_g)'}{N - G} \quad (4.2)$$

$$S_B = \frac{n \sum^G (\bar{Y}_g - \bar{Y})(\bar{Y}_g - \bar{Y})'}{G - 1} \quad (4.3)$$

$$S_T = \frac{\sum^G \sum^n (Y_{gk} - \bar{Y})(Y_{gk} - \bar{Y})'}{N - 1} \quad (4.4)$$

In the population, the covariance matrices within and between groups can also be aggregated as:

$$\Sigma_T = \Sigma_B + \Sigma_W \quad (4.5)$$

This decomposition is very interesting in order to analyse each component separately and can be also applied to our MTMM model (Equation 2.3). We are thus able to decompose the model in two parts. The subindices  $g$  and  $k$  are dropped for the sake of simplicity:

$$Y_{ij} = \underbrace{m_{Bij} M_{Bj} + t_{Bij} T_{Bi}}_{Y_{Bij}} + \underbrace{e_{Bij} + m_{wij} M_{wj} + t_{wij} T_{wi}}_{Y_{Wij}} + e_{wij} \quad (4.6)$$

Härnqvist (1978) proposes to do factor analysis on the within and between sample covariance matrices. Muthén (1989, 1990) shows that this can lead to biased estimates and suggests a maximum likelihood (ML) approach to estimate the population parameters of models of the CFA family by maximum likelihood on multilevel data structures.

If we have  $G$  balanced groups sizes (in our case egos) equal to  $n$  (in our case number of alters evaluated by each ego, thus the total sample size is  $N=Gn$ ) then  $S_w$  is the ML estimator of  $\Sigma_w$ , with sample size  $N-G$  and  $S_B$  is the ML estimator of  $\Sigma_w+c\Sigma_B$ , with sample size  $G-1$  with  $c$  equal to the common group size  $n$  (Hox, 1993). Then, for large samples the expected values are:

$$E(S_w) = \Sigma_w \tag{4.7}$$

$$E(S_B) = \Sigma_w + c\Sigma_B \tag{4.8}$$

where Equation 4.8 can be considered to be a multivariate equivalent to that encountered in one-way ANOVA with a random factor (e.g., Jackson and Brashers, 1994).

We can better understand Equations 4.6 to 4.8 in Figure 4, which is the two-level version of the path diagram in Figure 1.

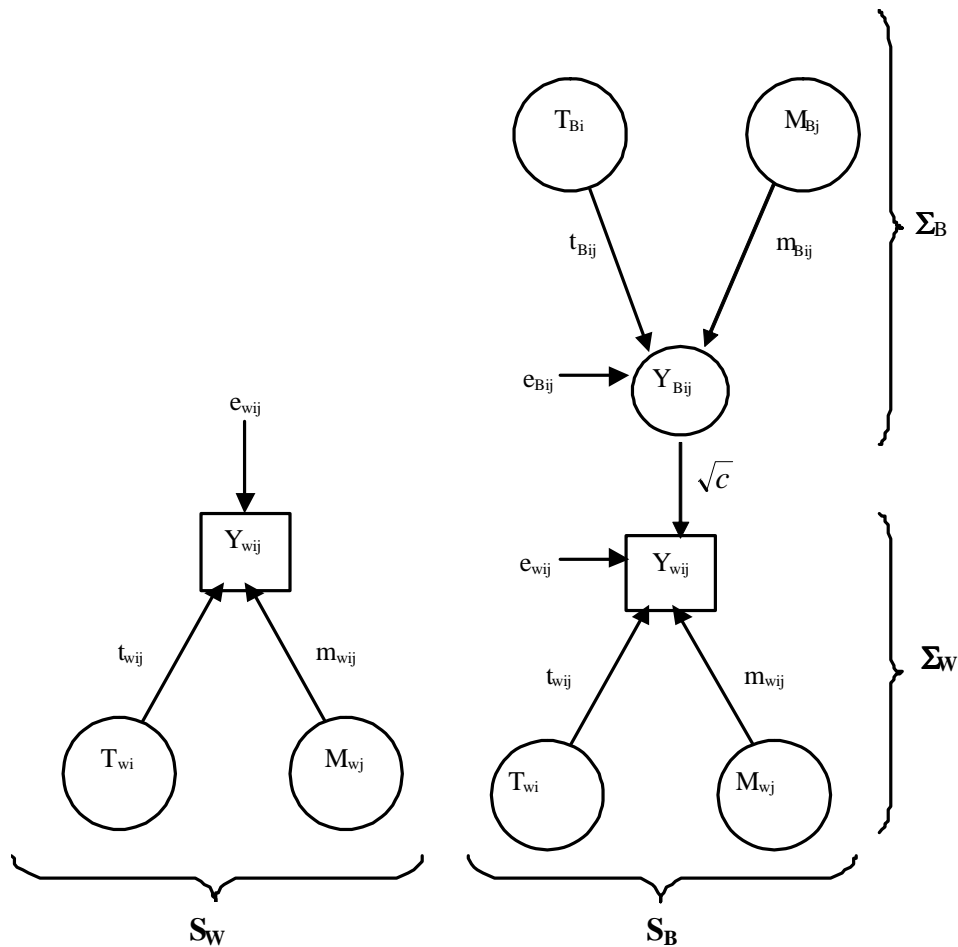


Figure 4: Multilevel CFA MTMM Model.

Note that  $\mathbf{S}_B$  estimates both the within structure ( $\Sigma_W$ ) and  $c$  times the between structure ( $\Sigma_B$ ) and is thus a biased estimate of  $\Sigma_B$ . This model can be estimated with standard structural equation modelling software if  $\mathbf{S}_W$  and  $\mathbf{S}_B$  are treated as two groups in a multiple group model with sample sizes  $N-G$  and  $G-1$  respectively. The variables in  $\mathbf{S}_W$  are only affected by the within factors, and the variables in  $\mathbf{S}_B$  are affected by both the within and between factors, weighted by a scaling factor  $\sqrt{c}$ . More recently developed software like Mplus (Muthén & Muthén, 2001) hides this complication from the user.

Until now we assumed that groups were of the same size (balanced case). In the unbalanced case the situation is more complex.  $\mathbf{S}_W$  continues to be a ML estimator for  $\Sigma_W$  and thus Equation 4.7 still holds. The difference is that the estimation of  $\Sigma_B$  is more complex because we need a different expression for each group size  $n_g$  (Hox, 1993):

$$E(\mathbf{S}_B) = \Sigma_W + n_g \Sigma_B \quad (4.9)$$

The Full Information Maximum Likelihood (FIML) estimator thus implies to specify a separate between group model for each distinct group size. This is computationally complex. Therefore Muthén (1989,1990) proposes to utilise another estimator known as the Partial Maximum Likelihood or Muthén Maximum Likelihood (MUML) estimator, called pseudobalanced solution, too. It's necessary to use a  $c^*$  scaling parameter, which is close to the mean group size.

$$c^* = \frac{N^2 - \sum n_g^2}{N(G-1)} \quad (4.10)$$

Whereas FIML is an exact ML estimator, MUML is only an approximation, but it should produce a good estimation given large sample sizes. MUML has been reported to perform well if the group sample size  $G$  (in our case the number of egos) is at least 100. Otherwise, standard errors and test statistics can be biased (Hox and Mass, 2001). Hox and Mass suggest that the number of groups  $G$  is more important for the quality of estimation than the total sample size  $N$ , especially for estimates of between group parameters.  $N$  will anyway be large enough in most practical applications. In this study, we use MUML as we have a large enough number of groups ( $G=314$ ).

## 4.2 Goodness of fit assessment

The evaluation of the goodness of fit of the model is a complex task for which many statistical tools are available (e.g., Bollen & Long, 1993; Batista-Foguet & Coenders, 2000). First of all, the estimates must be checked for admissibility (e.g.,

variances may not be negative, correlations may not be larger than one, etc.). A first goodness of fit measure is the  $\chi^2$  statistic to test the null hypothesis of no parameter omission, with its associated  $v$  number of degrees of freedom (d.f.) and  $p$ -value. The statistical power of this test varies with the sample size. If we have a large sample, the statistical test will almost certainly be significant. Thus, with large samples, we will always reject our model. Conversely, with a very small sample, the model will always be accepted, even if it fits rather badly. Thus, other useful fit measures that quantify the fit of the model have been suggested. Among them are the Compared Fit Index (CFI) of Bentler (1990), the Tucker and Lewis Index (TLI), also known as Non Normed Fit Index (NNFI) of Tucker and Lewis (1973), the Root Mean Square Error of Approximation (RMSEA) of Steiger (1990) and many others. Values of RMSEA below 0.050 (Browne & Cudeck, 1993) and values of TLI and CFI above 0.950 (Hu and Bentler, 1999) are usually considered acceptable. Recent research has shown the TLI to be independent of sample size and to adequately penalize complex models (Marsh et al. 1996). The RMSEA is also often reported due to its potential for hypothesis testing.

These goodness of fit indices are often only reported for the entire model, which includes both the fit in the within model and the between model. Hox (2002) suggests a specific strategy to evaluate the goodness of fit of a multilevel model in order to make it possible to identify whether the missfit comes from the between or within parts.

In order to evaluate the fit of the between part, a saturated model (with zero degrees of freedom and thus with a perfect fit) must be specified for the within part and the researchers' model (in our case a CFA MTMM model) for the between part. A goodness of fit measure such as RMSEA can be computed for the between part from this between part  $\chi^2_{B,MTMM}$  statistic, the associated degrees of freedom ( $v_{B,MTMM}$ ) and the sample size ( $G-1$ ) at the between part:

$$RMSEA_B = \sqrt{\frac{\chi^2_{B,MTMM} - v_{B,MTMM}}{(G-1)v_{B,MTMM}}} \quad (4.11)$$

Other goodness of fit measures like the TLI require the comparison of the  $\chi^2$  statistic with that of a model specifying zero covariances among all pairs of variables (independence model). Thus we would specify an independence model for the between part and a saturated model for the within part to obtain a  $\chi^2_{B,indep}$  statistic and its associated degrees of freedom  $v_{B,indep}$ . Thus, the  $TLI_B$  for the between part of the model would be:

$$TLI_B = \frac{\frac{\chi^2_{B,indep}}{v_{B,indep}} - \frac{\chi^2_{B,MTMM}}{v_{B,MTMM}}}{\frac{\chi^2_{B,indep}}{v_{B,indep}} - 1} \quad (4.12)$$

In a similar manner, we could obtain  $RMSEA_w$  and  $TLI_w$  by specifying two models with a saturated between part, taking into account that the within sample size is  $N-G$ .

### 4.3 Interpretation

In a multilevel context, the evaluation of measurement quality can be much enriched. Quite trivially, we can obtain two reliabilities and two validities for each trait-method combination, that is, between and within. The fact that groups are respondents and individuals are stimuli evaluated by them makes these reliabilities and validities interpretable in a somewhat different way from standard multilevel analysis.

The between reliabilities and validities can be computed from the parameters of the between part of the model and can be interpreted with respect to the quality of the measurement of the egocentered network as a whole (average values of the traits for each ego computed over all his/her alters).

The within reliabilities and validities can be computed from the parameters of the within part of the model and can be interpreted in a classic psychometric sense in which each subject is a separate unit of analysis and thus variance is defined across stimuli presented to the same subject, not across subjects (e.g., Lord, 1980).

Hox (2002) suggests that percentages of variance cannot only be computed in each part of the model separately. The fact that the between and within scores add to a total score as in Equation 4.6, makes it possible to compute percentages of variance in other attractive ways. In our case, if we decompose the variance according to Equation 4.6 we have:

$$\begin{aligned} \text{Var}(Y_{ij}) = & m_{ij}^2_w \text{Var}(M_{jW}) + m_{ij}^2_B \text{Var}(M_{jB}) + \\ & t_{ij}^2_w \text{Var}(T_{iW}) + t_{ij}^2_B \text{Var}(T_{iB}) \\ & \text{Var}(e_{ijw}) + \text{Var}(e_{ijB}) \end{aligned} \quad (4.13)$$

In this paper we suggest that each of the six components in Equation 4.13 can have its own interpretation:

- Between method variance corresponds to differences among respondents (egos) in the use of methods. Thus it is in complete agreement with the usual definition of method effect (e.g., Andrews, 1980).
- Within method variance corresponds to differences in the use of methods among alters evaluated by the same ego. At the moment we cannot interpret this source of variance. We would expect it to be very low in most cases.
- Between trait variance is the error-free variance corresponding to differences in the average levels of the egos.
- Within trait variance is the error-free variance corresponding to differences in the alter evaluations made by the same ego.

- Between error variance is the error variance associated to measurements of average levels of the egos. Thus, it is somehow systematic as it is constant for all alters within the ego (otherwise it would average to zero).
- Within error variance is not systematic in any way and thus truly corresponds to the definition of pure random measurement error.

Thus, from the decomposition in Equation 4.13, percentages like the following could be of interest and can easily be computed. One can:

- compute overall reliabilities and validities by aggregating all trait, method, and error components, thus similar results to a classic (not multilevel) analysis of  $\mathbf{S}_T$  would be obtained.
- compute overall percentages of within and between variance by aggregating all within components and all between components.
- do the former only with respect of error free variance, that is compute the percentage of between and within trait variance over the total trait variance.
- compute a percentage of pure random error variance (i.e. within error variance) over the total variance of the observed variables (grand total, i.e. including all 6 components). The percentage of total variance explained by the other 5 components can be computed in a similar way.

## 5 Results

### 5.1 Overview of the analyses performed

We are going to carry out four different analyses. The first three analyses will be of the traditional sort, analysing  $\mathbf{S}_T$ ,  $\mathbf{S}_W$  and  $\mathbf{S}_B$  separately with a standard (i.e. not multilevel) MTMM model. The last analysis will be a multilevel analysis, thus considering the within and between levels simultaneously:

- **Analysis 1a:** traditional analysis on  $\mathbf{S}_T$ . ML estimation.
- **Analysis 1b:** as 1a but using cluster sample formulae for the standard errors and goodness of fit indices (Muthén & Satorra, 1995). In fact, cluster samples are also an example of hierarchical data. Thus, even if we are only interested in the total scores, we can take the hierarchical structure of the data into account in this way. This procedure uses a mean-adjusted chi-square test statistic that is robust to non normality and to dependence among observations.
- **Analysis 2:** traditional analysis on  $\mathbf{S}_W$ , ML estimation.
- **Analysis 3:** traditional analysis on  $\mathbf{S}_B$ , which is a biased estimate of  $\Sigma_B$ . ML estimation. Analyses 2 and 3 together constitute the recommendation of Härnqvist (1978).

- **Analysis 4:** multilevel analysis, to fit  $\Sigma_W$  and  $\Sigma_B$  simultaneously. MUML estimation.

All analyses will be done using the Mplus2.12 program (Muthén & Muthén 2001). We will compare the traditional analyses (overall 1b, 2 and 3) to the multilevel analysis (4). Large differences are expected at least with analysis 3 which, if confirmed, will make the use of multilevel analysis (4) more advisable.

Some of the traditional MTMM analysis can be encountered in the social network literature. Hlebec (1999) and Ferligoj and Hlebec (1999) performed analysis 1a on complete network data and Kogovšek et al. (2002) analysis 3 on egocentered network data.

## 5.2 Goodness of fit of the models and specifications search

In Table 1, we can observe the goodness of fit of the different analyses ( $\chi^2$  statistic, TLI and RMSEA). The table also shows the changes in the specification that we have to make in order to obtain an admissible solution, as a few negative variances were obtained in the first specification, which had to be fixed at zero. Analysis 4, which includes both the within and between part, understandably required a larger number of respecifications.

**Table 1:** Goodness of fit statistics.

Analysis	1a ( $S_T$ ) ML	1b ( $S_T$ ) ML complex	2 ( $S_W$ ) ML	3 ( $S_B$ ) ML	4 ( $\Sigma_T$ and $\Sigma_W$ ) MUML
Initial					
$\chi^2$ statistic	191.074	140.579	112.210	82.068	149.313
d.f. (v)	15	15	15	15	30
TLI	0.959	0.958	0.971	0.930	0.967
RMSEA	0.093	0.078	0.078	0.119	0.054
Respecifications	var( $M_{2T}$ )=0	var( $M_{2T}$ )=0	var( $M_{2W}$ )=0	var( $M_{2B}$ )=0	t <sub>12b</sub> =1 var( $M_{1B}$ )=0 var( $M_{2W}$ ) = 0 var( $e_{41B}$ ) =0
$\chi^2$ statistic	192.993	141.925	112.401	82.644	185.333
d.f. (v)	16	16	16	16	34
TLI	0.961	0.961	0.973	0.934	0.963
RMSEA	0.090	0.076	0.075	0.115	0.057

A part from analysis 3, which yielded the worst goodness of fit, the goodness of fit of the final models of each analysis laid on the border between what can be considered a good or a bad fit. The model was rejected by the  $\chi^2$  statistic, RMSEA



was above (i.e. worse than) the commonly accepted threshold of 0.05 and TLI was above (i.e. better than) the threshold of 0.95.

However, the  $\chi^2$  statistic and hence the RMSEA may be somewhat inflated by the fact that data are ordinal (Babakus et al., 1987; Muthén and Kaplan, 1985) and by the fact that group sizes are unbalanced (Hox & Mass, 2001). Our group size distribution has a minimum of 1 alter per ego, a maximum of 13, a mean of 4.36, a standard deviation of 2.14 and a coefficient of variation of 0.49) The data simulated by Hox & Mass had a coefficient of variation equal to 0.50 and the  $\chi^2$  statistic was reported to have a positive bias of 8.6%. Removing the constraints in Equation 2.5 did not improve the fit (for instance if this is done on the final specification of analysis 4, the fit actually gets worse, as TLI=0.960 and RMSEA=0.059) and thus the constraints are maintained.

Analyses 1a and 1b report quite different goodness of fit measures, which suggests that it is important to use the corrections for complex samples when analysing  $S_T$  on hierarchical data.

**Table 2:** decomposition into 6 variance components. Analysis 4.

	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>
trait variance within				
M <sub>1</sub>	0.80	0.57	0.68	0.47
M <sub>2</sub>	0.77	0.55	0.65	0.45
method variance within*				
M <sub>1</sub>	0.03	0.03	0.03	0.03
M <sub>2</sub>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
error variance within				
M <sub>1</sub>	0.16	0.16	0.17	0.22
M <sub>2</sub>	0.14	0.13	0.13	0.17
trait variance between				
M <sub>1</sub>	0.17	0.06	0.10	0.13
M <sub>2</sub>	0.17	0.06	0.10	0.13
method variance between*				
M <sub>1</sub>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
M <sub>2</sub>	0.01	0.01	0.01	0.01
error variance between*				
M <sub>1</sub>	0.02	0.03	0.04	<b>0.00</b>
M <sub>2</sub>	0.04	0.02	0.02	0.07

\* Boldfaced for variances constrained to zero.

The goodness of fit of the final specification for analysis 4 can be decomposed into the within and the between part ( $\chi^2_B = 52.432$  with  $\nu_B = 18$  d.f.,  $TLI_B = 0.777$ ,  $RMSEA_B = 0.078$ ,  $\chi^2_W = 90.635$  with  $\nu_W = 16$  d.f.,  $TLI_W = 0.973$ ,  $RMSEA_W = 0.066$ ).

The fit thus seems to be worse for the between part of the model. For Analysis 4, all variances of traits, methods and errors were significantly different from zero except for the ones constrained in the specification search process ( $M_{1B}$ ,  $M_{2W}$ ,  $e_{41B}$ ) and  $e_{11B}$ . This suggests that trait, method and error variances operate both at the within and between levels and that none of the factors must be removed from the model specification.

Table 2 shows the variance decomposition according to Equation 4.13 for the eight variables (trait-method combinations) obtained from analysis 4. From this table, within, between and total reliabilities and validities and all other results described in Section 4.3 can be obtained. Boldfaced values are fixed to zero.

In the following subsections we show the final results of this analysis 4 in greater detail while we compare them to the traditional analyses. The first group of results are about  $\Sigma_W$  (thus involving analyses 2 and 4), the second group of results are about  $\Sigma_B$  (analyses 3 and 4) and the third group of results are about  $\Sigma_T$  (analyses 1 and 4). No distinction is made between analyses 1a and 1b as only goodness of fit measures change, not the point estimates.

### 5.3 Within part. Comparison of analyses 2 and 4

Table 3 presents the most commonly used estimates in an MTMM model, reliability and validity coefficients (square roots of Equations 2.6 and 2.7 respectively) and trait correlations, that is correlations corrected for measurement error.

According to Equation 4.7, the results of analyses 2 and 4 should be about the same. If we carefully study Table 3, we can confirm this equality: the results are virtually the same. Besides, both analyses required constraining the variance of  $M_2$  to zero in order to be admissible (see Table 1).

We find that closeness ( $T_2$ ) and importance ( $T_3$ ) are very highly correlated at the within level. This means that for a given ego, alters considered to be very close are also considered to be very important. Frequency of contact ( $T_1$ ) has moderate correlations with both abovementioned traits. Frequency of upsetting has lower correlations, but positive, thus meaning that being upset by an alter is not as negative as it may appear. Actually, the alters upsetting you the most are the ones you feel closest too, maybe because you contact them more often (actually the correlation between frequency of contact and frequency of upsetting is positive) or because you have higher expectations and thus can get upset by a lesser thing.

**Table 3:** Within part. Comparison of analyses 2 ( $S_w$ ) and 4 (multilevel).

	Analysis 2				Analysis 4			
	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>
Reliability coefficients								
M <sub>1</sub>	0.92	0.89	0.90	0.84	0.92	0.89	0.90	0.84
M <sub>2</sub>	0.92	0.90	0.91	0.85	0.92	0.90	0.91	0.85
Validity coefficients*								
M <sub>1</sub>	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97
M <sub>2</sub>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
Trait correlations								
T <sub>1</sub>	1.00				1.00			
T <sub>2</sub>	0.57	1.00			0.57	1.00		
T <sub>3</sub>	0.58	0.99	1.00		0.58	0.99	1.00	
T <sub>4</sub>	0.41	0.26	0.31	1.00	0.41	0.25	0.31	1.00

\* Boldfaced for variances constrained to zero.

As regards measurement quality at the within level, which is interpreted in a psychometric sense within a subject and across stimuli, Table 3 shows that frequency of contact (T<sub>1</sub>) is measured with the highest reliability and the frequency of being upset (T<sub>4</sub>) with the lowest for all methods. Telephone interviewing (M<sub>2</sub>) has higher reliability than personal interviewing (M<sub>1</sub>) for all traits. Validity coefficients referring to M<sub>2</sub> are equal to 1 because we have constrained the variance of this method to zero, since it was negative. This may simply mean that this variance was very low in the population, so that a negative sample estimate occurred just by chance; the estimate was indeed very low (about 1% of total within variance) and non-significant. Validity coefficients for M<sub>2</sub> are similar and high for all traits.

#### 5.4 Between part. Comparison of analyses 3 and 4

Table 4 presents reliability and validity coefficients and trait correlations for the between part, which can be obtained for analyses 3 and 4. Two insignificant negative method variances are constrained to zero, as shown in Table 1, and boldfaced.

If we compare both analyses we find very interesting results. According to Equation 4.8, the results should not be the same. If we study Table 4 carefully, we can confirm this inequality. The reliability coefficients are different in a rather

non-systematic way. The validity coefficients are not comparable, because different constraints are applied. The comparison of trait correlations is rather more interpretable. Equation 4.8 suggests that the analysis of  $\mathbf{S}_B$  is a combination of the within and between structures. Trait correlations obtained by analysis 3 are indeed half way between trait correlations within and between obtained by analysis 4. In any case, what Table 4 shows most clearly is that differences can be large, which suggest that an analysis of  $\mathbf{S}_B$  does badly at estimating the between structure of the data.

**Table 4:** Between part. Comparison of analyses 3 ( $\mathbf{S}_B$ ) and 4 (multilevel).

	Analysis 3				Analysis 4			
	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>
Reliability coefficients*								
M <sub>1</sub>	0.88	0.84	0.86	0.88	0.95	0.83	0.86	<b>1.00</b>
M <sub>2</sub>	0.97	0.91	0.93	0.88	0.91	0.88	0.90	0.82
Validity coefficients*								
M <sub>1</sub>	0.98	0.97	0.97	0.97	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
M <sub>2</sub>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.98	0.94	0.96	0.97
Trait correlations								
T <sub>1</sub>	1.00				1.00			
T <sub>2</sub>	0.23	1.00			-0.25	1.00		
T <sub>3</sub>	0.35	0.98	1.00		0.10	0.99	1.00	
T <sub>4</sub>	0.27	-0.03	0.07	1.00	0.16	-0.39	-0.17	1.00

\* Boldfaced for variances constrained to zero.

Given the large differences, for the remaining of this subsection we interpret the theoretically correct results of analysis 4. As regards the trait correlation matrix we again find that closeness (T<sub>2</sub>) and importance (T<sub>3</sub>) are very highly correlated. A more surprising finding is the very low correlation among all other pairs of traits, some of which are even negative. It must be taken into account that at the between level trait correlations refer to ego averages. For instance, at the between level it seems that egos with higher average frequency of contact (T<sub>1</sub>) do not feel more close (T<sub>2</sub>), on average, to their alters. On the contrary, at the within level, alters with whom one particular ego meets more frequently are the ones that particular ego feels closest to.

In Table 4 we are also able to observe reliability and validity coefficients at the between level, thus reflecting measurement quality of the ego averages across all alters. Unlike the case was at the within level, the telephone method (M<sub>2</sub>) is not always more reliable than the personal method. Validity coefficients of M<sub>2</sub> equal 1

for all traits, because we have constrained the variance of this method to zero. The validity coefficients for  $M_1$  are similar and high for all traits. In average over all variables, it cannot be said that measurement quality differs much from the within to the between level (the average reliability coefficients over all 8 variables are in fact equal up to the first two decimal places).

Kogovšek et al (2002) also analysed the  $S_B$  matrix for a somewhat different data set including a third method and two more samples of egos. Kogovšek et al (2002) reported  $M_2$  (telephone) to be more valid for all traits and to be more reliable for all traits but one. In spite of the differences in the sample used, this is much the same conclusion that can be drawn from analysis 3 in Table 4. Kogovšek (2002) argued that the telephone method may be more valid than face-to-face because it is more anonymous and more reliable because, being a faster means of communication, only the most important alters tend to be named. The finding of Kogovšek et al (2002) that the telephone mode produces good quality data is specially relevant in the social network literature because it contradicts the common finding in other research fields that the face-to-face mode produces data of better quality (e.g., Groves, 1989 and references therein). In our analyses it is not so clear that telephone method ( $M_2$ ) is better than face-to-face method ( $M_1$ ) at the between level. However, our analyses in the previous subsection did show that  $M_2$  produces better quality data at the within level, and an analysis of the  $S_B$  such as the one done by Kogovšek et al (2002) is inevitably contaminated by the within level structure according to Equation 4.8.

**Table 5:** Overall analysis. Comparison of analyses 1 ( $S_T$ ) and 4 (multilevel).

	Analysis 1				Analysis 4			
	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>
Reliability coefficients								
$M_1$	0.91	0.88	0.89	0.86	0.92	0.88	0.90	0.86
$M_2$	0.93	0.90	0.92	0.85	0.92	0.90	0.91	0.85
Validity coefficients*								
$M_1$	0.98	0.97	0.98	0.97	0.98	0.98	0.98	0.97
$M_2$	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	1.00	0.99	0.99	0.99
Trait correlations								
T <sub>1</sub>	1.00				1.00			
T <sub>2</sub>	0.46	1.00			0.46	1.00		
T <sub>3</sub>	0.50	0.99	1.00		0.50	0.99	1.00	
T <sub>4</sub>	0.36	0.15	0.22	1.00	0.36	0.16	0.23	1.00

\* Boldfaced for variances constrained to zero.

## 5.5 Overall analysis. Comparison of analyses 1 and 4

Table 5 presents overall reliability and validity coefficients, which can directly be obtained for analysis 1 and, by aggregating trait, method and error variances, also for analysis 4. One insignificant negative method variance is constrained to zero, as shown in Table 1, and boldfaced. Overall trait correlations for analysis 4 are computed by taking overall trait variances and covariances as the sum of between and within trait covariances, as in Equation 4.5.

If we compare both analyses we find very interesting results. According the theory explained before, the results should not be the same, but should be similar and comparable. If we study the tables carefully, we can confirm it. The trait correlations and reliability coefficients are very similar and even the validity coefficients are, in spite of the constraint of some variances to zero. The analysis of  $S_T$  may then be appropriate if one is only interested in overall parameter estimates, provided that correct test statistics are employed (i.e. analysis 1b).

If we consider the results of analysis 4, we are able to observe that the reliability and validity coefficients of both methods are quite similar, although  $M_2$  (telephone) is slightly better, except for the reliability when measuring  $T_4$  (frequency of being upset). Reliability coefficients are very high and validity coefficients are even more so, which is partly due to the fact that  $\text{var}(M_{1B})$  and  $\text{var}(M_{2W})$  have been constrained to zero, so that for  $M_1$  only the within level is counted and for  $M_2$  only the between level.

As suggested in section 4.3, when considering the overall model, the results of analysis 4 can be used to decompose variance in many interesting ways by combining interesting sets of the 6 variance components in Equation 4.13 and Table 2. Some of these results are shown in Table 6.

**Table 6:** Some interesting percentages of variance based on analysis 4.

	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>
$t_{ij}^2_w \text{Var}(T_{iW}) / [ t_{ij}^2_w \text{Var}(T_{iW}) + t_{ij}^2_B \text{Var}(T_{iB}) ]$	M <sub>1</sub> 0.82	0.90	0.87	0.79
	M <sub>2</sub> 0.82	0.90	0.87	0.78
$\text{Var}(e_{ijw}) / \text{Var}(Y_{ij})$	M <sub>1</sub> 0.13	0.19	0.16	0.26
	M <sub>2</sub> 0.12	0.17	0.14	0.20

The first part of Table 6 shows the percentage of within trait variance over all trait variance. The results show that most of the error free variance corresponds to the within level. This means egos really discriminate among different alters, which may also be an indicator of measurement quality. The second part of Table 6 shows the percentage of true random error variance (i. e. within error variance, as

argued in Section 4.3) over the total variance. One minus this percentage (or its square root) could be an alternative measure of reliability and would show measures with  $M_2$  (telephone) to be the most reliable and measures of  $T_4$  (frequency of upsetting the ego) the least.

## 6 Discussion

In this article we have used CFA for MTMM model, whose results are equivalent to TS model (Coenders & Saris, 2000) to study the quality of egocentered network data. We have used an additional constraint in the model that makes the  $t_{ij}$  trait loadings to be constant within method in order to increase the stability of the model. If we impose this assumption standard errors get much lower, which was quite valuable on a data set with only two methods.

As egocentered network data are hierarchical, we performed a multilevel MTMM analysis. Muthén's approach (1989, 1990, 1994) is used as we have a large enough number of groups. We compared the results of this multilevel analysis to those obtained using traditional analyses of the global, within and between covariance matrices. The traditional analysis of the between covariance matrix proved to yield misleading results, which leads to the recommendation to use the multilevel analysis, which provides much more detailed information and thus a much richer view on measurement quality from one single program run. However, if only the within data are of interest, a traditional analysis of the within covariance matrix could also be performed. In the same way, if only the overall data are of interest, an analysis of the overall covariance matrix is also possible, provided that appropriate corrections are made on standard errors and test statistics.

In our multilevel analysis, we can immediately obtain two reliabilities and validities for each trait-method combination, that is between and within egos. Each of them has a different interpretation. It is also possible compute overall reliabilities and validities by aggregating all trait, method and error components in order to obtain similar results to a classic (not multilevel) analysis of the overall covariances. As is usually done, we can also asses which percentage of variance is due to within and between differences. However, even more useful variance percentages can be obtained by combining different within and between components in a meaningful way (Hox, 2002) depending on the results one is interested in for a particular research problem.

We can also evaluate the goodness of fit of the multilevel model in such a way as to identify whether the missfit comes from the between or within parts of the model (Hox, 2002). In this paper we could thus find that the between part of the model fits worse.

According Kogovšek, (2002), telephone interviewing was more reliable and valid than the face-to-face method. According to de Leeuw (1992), the advantages

of telephone interviewing are larger for sensitive questions, category into which social support measures can be considered to fall. After our reanalysis of the same data, we conclude that is not so clear that telephone is more reliable than face to face. It depends on whether the within or the between level is considered. Telephone is better than face to face at the within level, and about equal to face to face at the between level. Differences in measurement quality can also be encountered for different traits as well. Frequency of contact is the most reliable trait in almost all cases, which could be so because mere frequency is easier for the respondent to interpret than traits involving feelings such as closeness, importance and upset.

## References

- [1] Althausen, R.P., Heberlein, T.A., and Scott, R.A. (1971): A Causal assessment of validity: The augmented multitrait-multimethod matrix. In H.M. Blalock, Jr. (Ed.): *Causal Models in the Social Sciences*. Chicago: Aldine, 151-169.
- [2] Alwin, D. (1974): An analytic comparison of four approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H.L. Costner (Ed.): *Sociological methodology 1973-1974*. San Francisco: Jossey-Bass, 79-105.
- [3] Andrews, F.M. (1984): Construct validity and error components of survey measures: a structural modeling approach. *Public Opinion Quarterly*, **48**, 409-442.
- [4] Babakus, E., Ferguson, C.E., and Jöreskog, K.G. (1987): The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, **24**, 222-229.
- [5] Bagozzi, R.P. and Yi, Y. (1991): Multitrait-multimethod matrices in consumer research. *Journal of Consumer Research*, **17**, 426-439.
- [6] Batista-Foguet, J.M. and Coenders, G. (2000): *Modelos de Ecuaciones Estructurales*. Madrid: La Muralla
- [7] Bentler, P.M. (1990): Comparative fit indexes in structural models. *Psychological Bulletin*, **107**, 238-246.
- [8] Bollen, K.A. and Long, J.S. (1993): *Testing Structural Equation Models*. Newbury Park, Ca.: Sage.
- [9] Brannick, M.T. and Spector, P.E. (1990): Estimation problems in the block-diagonal model of the multitrait-multimethod matrix. *Applied Psychological Measurement*, **14**, 325-339.



- [10] Browne, M.W. (1984): The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, **37**, 1-21.
- [11] Browne, M.W. (1985): *MUTMUM, Decomposition of Multitrait-Multimethod Matrices*. Pretoria, South Africa: University of South Africa, Department of Statistics.
- [12] Browne, M.W. and Cudeck, R. (1993): Alternative ways of assessing a model fit. In K.A. Bollen and J.S. Long (Eds.): *Testing Structural Equation Models*. Thousand Oaks, Ca: Sage, 136-162.
- [13] Campbell, D.T. and Fiske, D.W., (1959): Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**, 81-105.
- [14] Carmines, E.G. and Zeller, R.A. (1979): *Reliability and Validity Assessment*. Newbury Park, Ca.: Sage.
- [15] Coenders, G. and Saris, W.E. (2000): Testing nested additive, multiplicative, and general multitrait-multimethod models. *Structural Equation Modeling*, **7**, 219-250.
- [16] Ferligoj, A. and Hlebec, V. (1999): Evaluation of social network measurement instruments. *Social networks*, **21**, 111-130.
- [17] Härnqvist, K (1978): Primary mental abilities at collective and individual levels. *Journal of Educational Psychology*, **70**, 706-716.
- [18] Heise, D.R. (1969): Separating reliability and stability in test-retest correlations. *American Sociological Review*, **34**, 93-101.
- [19] Hlebec, V. (1999): *Evaluation of Survey Measurement Instruments for Measuring Social Networks*, Doctoral dissertation, University of Ljubljana, Slovenia.
- [20] Hox, J.J. (1993): Factor analysis of multilevel data: gauging the Muthén method. In J.H.L. Oud and R.A.W. van Blokland-Vogeleang (Eds.): *Advances in Longitudinal and Multivariate Analysis in the Behavioral Sciences*. Nijmegen, the Netherlands, ITS,141-156.
- [21] Hox, J.J. (2002): *Multilevel Analysis. Techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum.
- [22] Hox, J.J. and Mass, C.J.M. (2001): The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, **8**, 157-174
- [23] Hu, L. and Bentler, P.M. (1999): Cutoff criteria for fit indices in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1-55
- [24] Jackson, S. and Brashers, D.E. (1994): *Random Effects in ANOVA*. Newbury Park, Ca.: Sage.

- [25] Groves, R.M. (1989): *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- [26] Kenny, D.A. and Kashy, D. A. (1992): Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, **112**, 165-172.
- [27] Kogovšek, T., Ferligoj, A., Coenders, G., and Saris, W.E. (2002): Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. *Social Networks*, **24**, 1-20.
- [28] De Leeuw, E.D. (1992): *Data quality in mail, telephone and face-to-face surveys*. Amsterdam: TT-Publikaties.
- [29] Lord, F.M. (1980): *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- [30] Lord, F. M. and Novick, M. R. (1968): *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- [31] Marsh, H.W. (1989): Confirmatory factor analysis of multitrait-multimethod data: Many problems and few solutions. *Applied Psychological Measurement*, **13**, 335-361.
- [32] Marsh, H.W. and Bailey, M. (1991): Confirmatory factor analyses of multitrait-multimethod data: Comparison of the behavior of alternative models. *Applied Psychological Measurement*, **15**, 47-70.
- [33] Marsh, H.W., Balla, J.R., and Hau, K.T. (1996): An evaluation of incremental fit indices: a clarification of mathematical and empirical properties. In G.A. Marcoulides and R.E. Schumacker (Eds.): *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah, NJ: Lawrence Erlbaum, 315-353.
- [34] Muthén, B. (1989): Latent variable modelling in heterogeneous populations. *Psychometrika*, **54**, 557-585.
- [35] Muthén, B. (1990): *Means and Covariance Structure Analysis of Hierarchical Data*. Los Angeles: UCLA. Statistics series, 62.
- [36] Muthén, B. (1994): Multilevel covariance structure analysis. *Sociological Methods & Research*, **22**, 376-398.
- [37] Muthén, B. and Kaplan, D. (1985): A Comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, **38**, 171-189.
- [38] Muthén, L.K. and Muthén, B. (2001): *Mplus User's Guide*. Los Angeles, Ca: Muthén & Muthén.
- [39] Muthén, B. and Satorra, A. (1995): Complex sample data in structural equation modeling. In P.V. Marsden (Ed.): *Sociological Methodology*. Washington D.C.: The American Sociological Association, 267-316.
- [40] Rindskopf, D. (1984): Structural equation models: empirical identification, Heywood cases and related problems. *Sociological Methods and Research*, **13**, 109-119.

- [41] Saris, W.E. (1990a): Models for evaluation of measurement instruments. In W.E. Saris and A. van Meurs (Eds.): *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait Multimethod Studies*. Amsterdam: North Holland, 52-80.
- [42] Saris, W.E. (1990b): The choice of a model for evaluation of measurement instruments. In W.E. Saris and A. van Meurs (Eds.): *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait Multimethod Studies*. Amsterdam: North Holland, 118-129.
- [43] Saris, W.E. and Andrews, F.M. (1991): Evaluation of measurement instruments using a structural modeling approach. In: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (Eds.): *Measurement Errors in Surveys*. New York: Wiley, 575-597.
- [44] Saris, W.E. (1995): Designs of models for quality assessment of survey measures. In W.E. Saris and Á. Münnich (Eds.): *The Multitrait Multimethod Approach to Evaluate Measurement Instruments*. Budapest: Eötvös University Press, 9-37.
- [45] Sherpenzeel, A. (1995): *A Question of Quality. Evaluation of Survey questions by Multitrait-Multimethod Studies*. Doctoral dissertation, University of Amsterdam. Leidschendam. The Netherlands.
- [46] Schmitt, N. and Stults, D.N. (1986): Methodology review. Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, **10**, 1-22.
- [47] Steiger, J.H. (1990): Structural model evaluation and modification. An interval estimation approach. *Multivariate Behavioural Research*, **25**, 173-180.
- [48] Tucker, L.R. and Lewis, C. (1973): A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, **38**, 1-10.
- [49] Werts, C.E. and Linn, R.L. (1970): Path analysis. Psychological examples. *Psychological Bulletin*, **74**, 193-212.