Measurement Quality in Indicators of Compositions.
A Compositional Multitrait-Multimethod Approach

Germà Coenders
Department of Economics, University of Girona
Valentina Hlebec
Faculty of Social Sciences, University of Ljubljana
Tina Kogovšek
Faculty of Arts and Faculty of Social Sciences, University of Ljubljana

Seminari d'Estadística de Girona
dijous 20 d'octubre 2011

Introduction

Statistical compositions consist of positive data arrays with a fixed sum. The commonest examples are proportions or percentages of the set of components of a total, whose sum can only be 1 or 100%. Compositional data are thus severely constrained.

Composition indicators are frequent in social science data collected by surveys:

- Budget surveys: percentages spent on the given good or service categories.
- Time-use surveys: the total amount of available time is usually constant (e.g. 24h).
- Compositional indicators in network surveys (e.g. % of family members, friends, …).

Compositional data do not lend themselves easily to standard statistical analyses:

- On the one hand, specialised techniques for compositional data are starting to appear (e.g. Thió-Henestrosa & Martín-Fernández, 2005).
- On the other hand, compositional data can be transformed so that they can be subject to standard statistical techniques as they are, or with minor modifications (Aitchison, 1986).

When it comes to assessing measurement quality of questions, standard statistical techniques such as confirmatory factor analysis (CFA) are commonly understood by social scientists, and the approach of transforming the data and keeping analyses standard shows greater promise.

Correlated uniqueness model for multitrait-multimethod (MTMM) designs

MTMM designs (Campbell & Fiske, 1959) are a well established approach to assess measurement quality of survey questions (Saris & Gallhofer, 2007). These designs consist of multiple measures of at least three factors (traits) with the same set of at least three measurement procedures (methods). So, these designs include *DM* measures, that is the number of methods (M) times the number of traits (D).

MTMM designs are usually analysed by means of CFA models, a particular case of structural equation models (SEM). A number of CFA models for MTMM data have been formulated and tested in the literature. Coenders and Saris (2000) showed the great flexibility of the so-called correlated uniqueness (CU) model (Marsh, 1989), of which many other MTMM models constitute particular cases. The CU model is a CFA model specified as follows.

Let x_{idm} be the measurement of individual i, for trait d with method m:

$$x_{idm}=\tau_{dm}+\lambda_{dm}t_{id}+e_{idm}$$

where t_{id} is the latent variable score of individual i corresponding to trait d and e_{idm} is the measurement error term of individual i, for trait d with method m. Each factor is represented by a component t_{id} and measured by M methods: $x_{id1}, x_{id2},\ldots, x_{idM}$

with the assumptions:

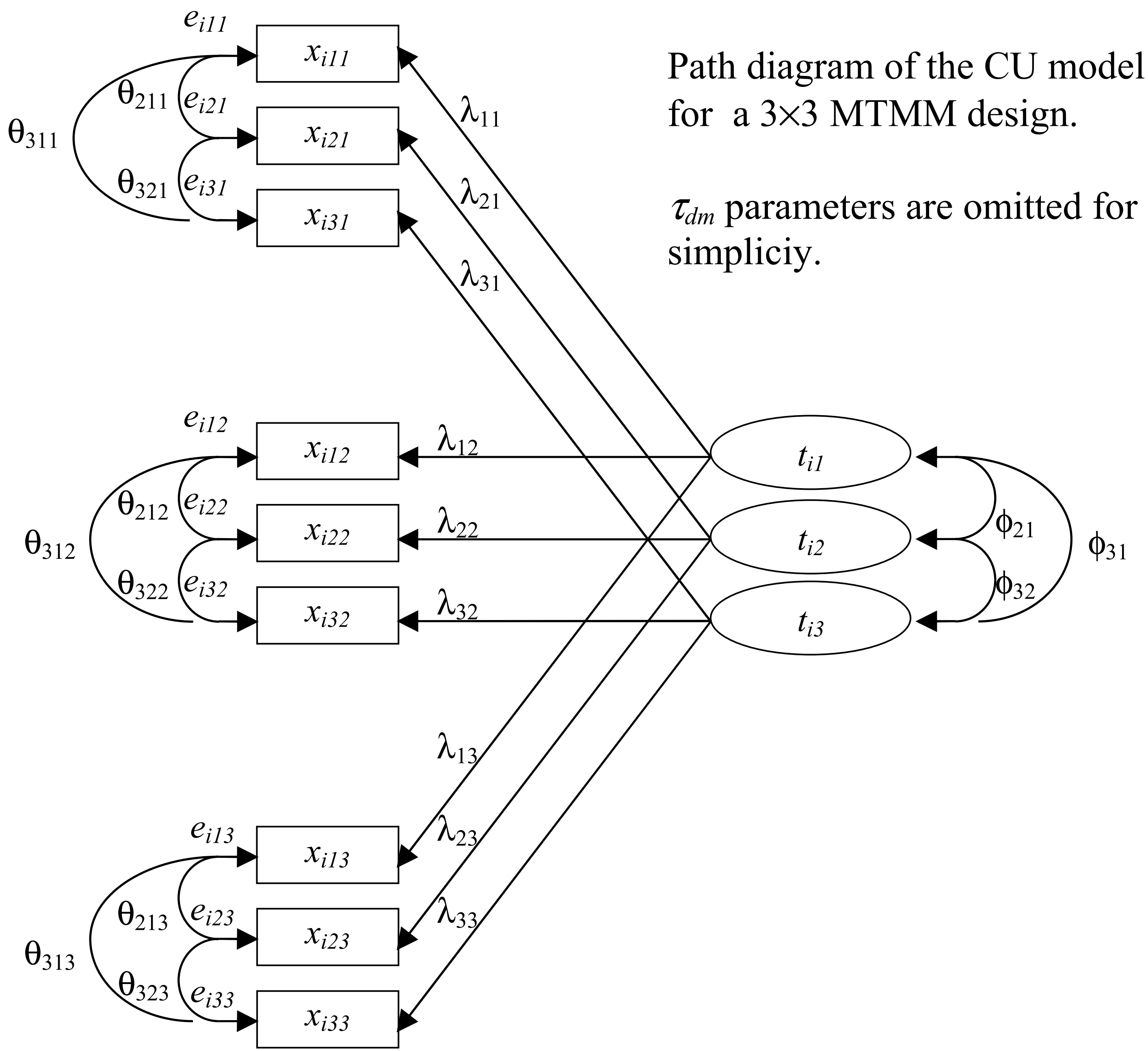
$$\mathrm{E}(t_{id})=\mathrm{E}(e_{idm})=0$$
$$\mathrm{cov}(e_{idm},e_{id'm})=\theta_{dd'm}$$
$$\mathrm{cov}(t_{id},t_{id'})=\phi_{dd'}$$
$$\mathrm{cov}(e_{idm},e_{id'm'})=0$$

The model parameters are:

- τ_{dm}: expected value of x_{idm}.
- λ_{dm}: factor loading of x_{idm} on trait t_{id}. It relates the scales of x_{idm} and t_{id}. One loading for each trait (i.e. when m=1) has to be constrained to unity for latent variable identification purposes (λ_{d1}=1).
- θ_{dm}: measurement error variance of e_{idm}.
- $\theta_{dd'm}$: covariance between two measurement error terms sharing a common method e_{idm} and $e_{id'm}$. In an MTMM design it is expected that the use of the same method involves common errors. These covariances are called method effects for this reason.
- ϕ_{dd}: variance of the trait latent variable t_{id}.
- $\phi_{dd'}$: covariance between two trait latent variables t_{id} and $t_{id'}$.

Two main measurement quality indicators can be obtained by the model:

- Standardized trait loadings λ_{dm} measure the strength of the relationship between observed scores and trait latent scores. Other measures of measurement quality can be obtained by re-expressing the standardized trait loadings. The squared standardized loading is the percentage of variance of x_{idm} explained by t_{id}. The standardized error variance is one minus the squared standardized loading. Of course, these sets of measures are mutually redundant and just one of them is enough.
- Intercepts τ_{dm}, $\tau_{dm'}$,… measure relative bias of several methods *m, m'*,… when measuring trait *d*. If $\tau_{dm}=\tau_{dm'}$, then there is no difference in the biases of *m* and *m'* when measuring trait *d*. If $\tau_{dm}>\tau_{dm'}$, then *m* yields systematically larger scores than *m'*.



Path diagram of the CU model for a 3×3 MTMM design.

τ_{dm} parameters are omitted for simpliciy.

Challenges in the analysis of compositional data

Compositional data concern the relative size of *D* components within a total, usually proportions over 1 or 100%: for instance % of friends, family, etc. in a personal network.

The study of the measurement quality of compositional data cannot be undertaken by just fitting the proportions or percentages to a SEM (e.g. to a CU model).

Unlike unconstrained data (e.g. number of friends, family… in the network), compositional data lie in a constrained space. A *D*-term composition measured on individual *i* with method *m* is:

$$x_{i1m}, x_{i2m}, \ldots, x_{iDm}$$

with the constraints:

$$0 \leq x_{idm} \leq 1 \text{ and } \sum_{d=1}^{D} x_{idm} = 1$$

The unconstrained data in absolute terms, which are often unknown, are:

$$s_{im}x_{i1m}, s_{im}x_{i2m}, \ldots, s_{im}x_{iDm}$$

where s_{im} is size for individual *i* as given with method *m*.

Aitchison (1986) warns against the problems of standard statistical tools on compositions:

- Compositional data are non-normal, ranging within 0 and 1, and often highly skewed.
- Compositional data have a constrained sum: one component can only increase if some other(s) decrease. This results in spurious negative correlations among components:

$$\sum_{d<d'} \operatorname{cov}(x_{idm}, x_{id'm}) = -(1/2)\sum_{d} \operatorname{var}(x_{idm})$$

- The true dimensionality of a set of compositional variables measured with a given method *m* is *D*-1. Analysis of all *D* dimensions leads to non-positive definite covariance matrices, perfect collinearity and the like.

In the context of SEM, constant sum data are referred to as ipsative data (Chan, 2003).

- Zero sum data are called additive ipsative data .
- Unit sum data (compositional data) are called multiplicative ipsative data.

While additive ipsative data have successfully been dealt with in the SEM context (Chan, 2003; Cheung, 2004), this is not the case for multiplicative ipsative data.

The problems reported by Aitchison (1986) also apply to the CU model with an important addition. Even if the absolute data $s_{im}x_{i1m}$ fit a CU model, the compositional x_{idm} data do not. Let us consider the model for D=4 components measured with method 1:

$$x_{i11}=\tau_{11}+\lambda_{11}t_{i1}+e_{i11}$$
$$x_{i21}=\tau_{21}+\lambda_{21}t_{i2}+e_{i21}$$
$$x_{i31}=\tau_{31}+\lambda_{31}t_{i3}+e_{i31}$$
$$x_{i41}=\tau_{41}+\lambda_{41}t_{i4}+e_{i41}$$

Both the expected and the individual compositions must add up to 1:

$$\tau_{11}+\tau_{21}+\tau_{31}+\tau_{41}=1$$
$$x_{i11}+x_{i21}+x_{i31}+x_{i41}=1$$

$$x_{i11}=\tau_{11}+\lambda_{11}t_{i1}+e_{i11}=1-\tau_{21}-\lambda_{21}t_{i2}-e_{i21}-\tau_{31}-\lambda_{31}t_{i3}-e_{i31}-\tau_{41}-\lambda_{41}t_{i4}-e_{i41}$$
$$e_{i11}=-\lambda_{11}t_{i1}-\lambda_{21}t_{i2}-e_{i21}-\lambda_{31}t_{i3}-e_{i31}-\lambda_{41}t_{i4}-e_{i41}$$

Any error is dependent on all traits and on all remaining errors, within a method. Given the true compositions t_{idm}, the observed component x_{i1m} can only increase if some others decrease. The CU model assuming each variable to load only on a trait is miss-specified.

Fortunately, the CU model for MTMM data is less problematic than standard factor analysis (FA) of compositional data. The CU model differs from many standard FA applications in the following:

- In the CU model each factor is a single component. In standard FA models each factor is a cluster of correlated components.
- In the CU model, the indicators of each factor are independent measurements of the single component. In standard FA models the indicators of each factor are the same measurements of the correlated components.
- The key correlations in the CU model are those among independent measurements of the same component. The key correlations in the standard CFA model are those among the same measurements of different components. The key correlations in the CU model are not spurious.

Compositional data transformations

The analysis of compositional data with standard statistical methods is possible after some kind of ratio transformation has been applied. Several ratio transformations have been suggested in the literature. Among them are the additive logratio transformation (alr), the centred logratio transformation (clr), both suggested by Aitchison (1986) and the isometric logratio transformation (ilr) suggested by Egozcue et al. (2003).

- The clr transformation leads to additive ipsative data and, thus, there is not much to be gained from it in this context.
- The ilr transformation is less intuitive and more complicated to compute but has appealing properties preserving distances among cases. It is thus appealing for such applications as cluster analysis or graphical displays.

The alr transformation is by far the easiest to compute and to interpret:

$$y_{idm}=\ln(x_{idm}/x_{iDm})=\ln(x_{idm})-\ln(x_{iDm})=\ln(s_{im}x_{idm})-\ln(s_{im}x_{iDm})$$
$$\text{with } d=1,2,...,D-1$$

- The alr transformed composition has one fewer dimension than the original composition.
- The alr transformed y_{idm} variables recover the full $-\infty$ to ∞ range. Whether the alr data follow a normal distribution or not will, of course, depend on the particular case at hand.

The CU model is simply estimated on the alr y_{idm} data on the $(D-1)M$-dimensional data set with standard methods for SEM estimation.

The CU model on alr data still has some limitations regarding parameter interpretation:

- Trait correlations tend to be positive because alr data have a common denominator. The correlations among ratios cannot be interpreted as correlations among the original absolute data $s_{im}x_{idm}$. The covariance between any two components contains the variance of the Dth component, which can only be positive:

$$\mathrm{cov}(y_{idm}, y_{id'm})=\mathrm{cov}(\ln(s_{im}x_{idm}),\ln(s_{im}x_{id'm}))+\mathrm{var}(\ln(s_{im}x_{iDm}))$$
$$-\mathrm{cov}(\ln(s_{im}x_{idm}),\ln(s_{im}x_{iDm}))-\mathrm{cov}(\ln(s_{im}x_{id'm}),\ln(s_{im}x_{iDm}))$$

- Likewise, error term covariances $\theta_{dd'm}$ are spurious and positive, as they contain the variance of the Dth error term. Therefore, they cannot be interpreted as method effects or measurement invalidity. The CU model is appropriate for compositional data because it includes error covariance parameters for all pairs of components measured with a given method. These error covariances play a methodological role and are not interpreted.

The main parameters of interest are thus standardized trait loadings (indicating measurement quality) and raw (i.e. non-standardized) intercepts (indicating relative bias).

Dealing with zero components

If either x_{idm} or x_{iDm} equal zero, y_{idm} cannot be computed.

An obvious first procedure is to amalgamate small components with many zeroes into larger ones. This is feasible if the amalgamated components have a degree of theoretical similarity.

In certain instances, some zero components result from individual characteristics. For instance, people who have never been employed cannot have co-workers in their social network (essential zeroes). When external variables are available to identify them, it may be advisable to narrow the definition of the target population and remove them from the sample.

After amalgamation of components and redefinition of the population, the remaining few zeroes may be understood as components which are too small to be detected:

- In a time-use diary with half hour intervals, a small amount of time devoted to an activity will likely not be recorded and components smaller than 1/48 will not be detected.
- In a social network questionnaire in which respondents are allowed to mention up to s_m members, components smaller than $1/s_m$ will not be detected.

Both examples differ in one respect. The first is based on numeric variables (time units) and the second on multinomial variables (counts of members with given characteristics).

In both cases, zeroes are substituted by a small amount which is likely to be undetected.

- For numeric variables, we define the smallest detectable proportion as δ_{idm}. Martín-Fernández et al. (2003) suggest replacing $x_{idm}=0$ with:

$$x'_{idm}=k\delta_{idm} \text{ with } 0<k<1.$$

  The authors suggest using $k=0.65$, and doing a sensitivity analysis on the choice of k.

- For multinomial variables, Pierotti et al. (2009)'s Bayesian approach replaces $x_{idm}=0$ with:

$$x'_{idm}=\frac{1}{D(s_{im}+1)}$$

Non-zero x_{idm} values have to be reduced in order to preserve the unit sum. As in Martín-Fernández et al. (2003) both in the numeric and multinomial case, $x_{idm}>0$ are replaced with:

$$x'_{idm}=x_{idm}\left(1-\sum_{x_{idm}=0}x'_{idm}\right)$$

Illustration. Data

The focus of this example are indicators of network composition obtained from egocentered networks: relationships between a single ego and a set of alters.

Once the names of alters are obtained with the so-called name generator questions (e.g. to whom would you ask for help if you would need…?), several additional questions (name interpreters) are posed to find out about characteristics of network members (age, gender) and ties connecting ego to her/his alters (type of relation, frequency of contacts, geographical distance…). The characteristics measured through name interpreters can be used to classify network members into components, for instance, percentages of partner, kin, friends and other members within the network.

Very often data about egocentered networks are collected in a survey setting. Several studies have addressed measurement quality of egocentered networks measured with surveys (e.g. Kogovšek 2006; Kogovšek et al., 2002; Kogovšek & Ferligoj, 2003; 2004; 2005; Lozar-Manfreda et al., 2004; Vehovar et al., 2008).

Our example uses the same data as Kogovšek et al. (2002). The data were collected in 2000 by computer-assisted telephone interview (CATI) and computer-assisted personal interview (CAPI) for a representative sample of 1033 inhabitants of Ljubljana, Slovenia.

The components (traits) used in this example are the percentages of network members represented by:

- 1: partner
- 2: friends
- 3: others
- 4: family (reference component for the alr transformation)

Kogovšek et al. (2002) use the telephone and face-to-face data collection modes, and two ways of ordering name interpreter questions “by alters” and “by questions”. The first, “by alters”, is to take each alter and to ask all name interpreter questions about him/her before moving to the next alter. The second way, “by questions”, is to take each question (e.g. on the alter’s relationship to ego) and ask this question for all alters before moving to the next question. The three different methods used are:

- 1: Face to face by alters
- 2: Telephone by questions
- 3: Telephone by alters

We expected that the quality of indices of composition should be highest when using the face-to-face data collection mode. This expectation derives from studies comparing the two data collection modes, which conclude that, for cognitively demanding questions, face-to-face interviews are preferred (Kogovšek et al., 2002; Kogovšek & Ferligoj, 2004; 2005).

Illustration. Results

Descriptive statistics of the imputed x' scores and the additive log ratio y scores

	Min	Max	Mean	St.dev.	Skewness	Kurtosis
x'_{11}	.013	.875	.124	.116	1.95*	6.92*
x'_{21}	.015	.938	.402	.235	0.09	-0.82*
x'_{31}	.015	.917	.145	.170	1.95*	3.96*
x'_{41}	.015	.906	.329	.210	0.42*	-0.38*
x'_{12}	.010	.458	.111	.093	1.13*	0.94*
x'_{22}	.016	.946	.405	.237	0.05	-0.83*
x'_{32}	.015	.917	.143	.167	1.88*	3.55*
x'_{42}	.013	.917	.340	.215	0.43*	-0.37*
x'_{13}	.010	.850	.112	.099	1.65*	5.00*
x'_{23}	.018	.953	.415	.223	-0.02	-0.64*
x'_{33}	.010	.850	.136	.148	1.65*	2.53*
x'_{43}	.013	.938	.338	.204	0.53*	-0.09

	Min	Max	Mean	St.dev.	Skewness	Kurtosis
y_{11}	-3.84	3.04	-1.07	1.40	0.49*	-0.07
y_{21}	-3.77	4.12	0.21	1.60	-0.11	-0.06
y_{31}	-3.84	3.50	-1.08	1.60	0.72*	-0.05
y_{12}	-3.62	2.76	-1.17	1.29	0.39*	-0.24
y_{22}	-3.60	4.11	0.18	1.66	-0.09	-0.17
y_{32}	-3.78	3.50	-1.13	1.56	0.64*	-0.30
y_{13}	-3.81	2.83	-1.22	1.26	0.41*	-0.18
y_{23}	-3.83	4.11	0.22	1.53	-0.22*	0.28
y_{33}	-3.83	4.02	-1.18	1.49	0.62*	-0.12

First subindex shows trait
(1: partner; 2: friends; 3 others; 4: family).
Second subindex shows method
(1: Face to face by alters; 2: Telephone by questions; 3: Telephone by alters).
* Significant skewness or kurtosis (α=5%).

The *x'* scores show roughly similar means for the different methods. All methods give the partner as the smallest component, others as the second to smallest, family as the second to largest and friends as the largest. Nearly all components have significant skewness and kurtosis. The smallest components have rather extreme coefficients.

The *y* scores are relative to the 4^{th} component (family). The mean values show friends to be a somewhat larger component than family and partner and others to be much smaller components. The degree of non-normality is much reduced.

The shaded cells show correlations between the same trait using two methods. These correlations are highest between methods 1 and 2 and lowest between methods 2 and 3. The diagonal sub-matrices show correlations among different traits using a common method and are all positive, as is often the case with alr scores.

Correlation matrix

	y_{11}	y_{21}	y_{31}	y_{12}	y_{22}	y_{32}	y_{13}	y_{23}	y_{33}
y_{11}	1.000								
y_{21}	.504	1.000							
y_{31}	.519	.453	1.000						
y_{12}	.738	.382	.382	1.000					
y_{22}	.364	.634	.384	.564	1.000				
y_{32}	.332	.413	.586	.479	.474	1.000			
y_{13}	.651	.321	.282	.615	.357	.314	1.000		
y_{23}	.277	.626	.320	.340	.634	.376	.497	1.000	
y_{33}	.342	.339	.578	.307	.337	.553	.462	.489	1.000

First subindex shows trait
(1: partner; 2: friends; 3 others).
Second subindex shows method
(1: Face to face by alters; 2: Telephone by questions; 3: Telephone by alters).

The CU model yielded a robust Yuan and Bentler χ^2 statistic 14.95 with 15 d.f. and *p*-value=0.455. The 90 Percent C.I. for RMSEA (Root Mean Square Error of Approximation) was 0.000 to 0.029. Other usual goodness of fit measures also revealed an excellent fit. CFI (Comparative Fit Index) and TLI (Tucker and Lewis Index) were both 1.000.

Measurement quality (standardized λ_{dm} trait loadings), is highest for method 1 (face to face by alters) for all three components. The second and third methods behave about equally well. Differences among methods are not dramatic when compared to sampling variability (confidence intervals). We expected that data quality of indices of composition should be highest when using the face-to-face data collection mode. The question order by questions or by alters is a side issue.

The partner to family ratio (trait 1) has the highest measurement quality. The partner is the single most prominent provider of social support. The ratio involving other network members (trait 3) has the lowest measurement quality. This category likely comprises the respondents' weakest ties.

As regards the expected τ_{dm} ratios, the largest differences are encountered in the ratio of the partner network to the family network (trait 1). Method 1 gives the largest ratio, method 2 being close to method 3. Partner was listed at the beginning of the response options list, which may have caused a primacy effect for the face-to-face mode.

Measurement quality estimates and 95% C.I. from the CU model

	Standardized λ_{dm}: loadings			τ_{dm}: expected values		
	lower 95% limit	estimate	upper 95% limit	lower 95% limit	estimate	upper 95% limit
y_{11}	.811	.883	.955	-1.144	-1.047	-0.950
y_{21}	.757	.829	.901	0.103	0.216	0.330
y_{31}	.710	.786	.862	-1.188	-1.072	-0.957
y_{12}	.745	.816	.886	-1.265	-1.177	-1.090
y_{22}	.698	.771	.844	0.078	0.190	0.303
y_{32}	.671	.750	.830	-1.242	-1.135	-1.029
y_{13}	.687	.751	.815	-1.307	-1.221	-1.136
y_{23}	.716	.782	.847	0.099	0.203	0.308
y_{33}	.667	.737	.807	-1.290	-1.186	-1.083

First subindex shows trait
(1: partner; 2: friends; 3 others).
Second subindex shows method
(1: Face to face by alters; 2: Telephone by questions; 3: Telephone by alters).

Discussion

Indicators of network composition measured with surveys have not yet been properly evaluated with MTMM models owing to the difficulties involved by the compositional nature of these indicators. In this article we highlight the necessary data transformations, the appropriate type of MTMM model and the required changes in the interpretation of model parameters. Whenever possible, the simplest methodological choices have been made, in order to narrow the gap between methodology development and application.

In this example we have used data obtained by name generator questions which did not limit the number of mentioned alters. The approach described in this article is even more appropriate for network compositions with constrained name generator questions (imposing a certain number of alters for all respondents). Under these formats, total network size becomes constant and thus the researcher has no choice but to use compositional data approaches.

Further research in the area will include:

- Extending the CU model to a full SEM in which components are predicted by a set of covariates or in which components are covariates predicting a set of outcome variables.
- Evaluating the relative merits of other more complex data transformations such as the ilr.
- Meta analyzing measurement quality of compositional data obtained in several studies.
- Including web data collection.