Available online at www.sciencedirect.com
SCIENCE DIRECT®
Social Networks 27 (2005) 205–229
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Effects on reliability and validity of egocentered network measurements

Tina Kogovšek*, Anuška Ferligoj

Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia

Abstract

This paper examines the reliability and validity of egocentered networks. Reliability and validity are estimated by the multitrait-multimethod (MTMM) approach. A split ballot MTMM design [Saris, W.E., 1999. Forced choice or agree/disagree questions? An evaluation by the split ballot MTMM experiment. In: Proceeding of the Meeting of the IRMCS, pp. 122–146; Kogovšek, T., Ferligoj, A., Coenders, G. 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] is used, in which separate groups of respondents received different combinations of two methods. The effect of factors such as the methods used and the personal characteristics of respondents that can affect the quality of data was estimated by a meta analysis.

Measurement method, type of question, network size, age, gender, extraversion and emotional stability all had statistically significant effects on the validity of measurement. After the list of alters is obtained with name generators, name interpreter questions can be asked in two ways. One way ("by alters") is to take each alter individually and to ask all the questions about him/her, going alter by alter until the end of the list of alters. The other way ("by questions") is to take the question and ask this question for all alters on the list, going question by question until the end of the list of name interpreter questions. Telephone interviewing (both by alters and by questions) gave more valid measurements than face-to-face interviews.

Behavioral questions were more valid than questions with emotional content. The characteristics of ties were more validly measured in smaller networks. With reference to respondents' personal characteristics younger respondents, men, extraverted and emotionally stable respondents all had more valid measurements. Reliability was significantly affected by the measurement method, the type of question and age. The telephone/by alters method was the most reliable measurement method.

* Corresponding author. Tel.: +386 1 58 05 285; fax: +386 1 58 05 101.
E-mail address: Tina.Kogovsek@guest.arnes.si (T. Kogovšek).

0378-8733/$ – see front matter
doi:10.1016/j.socnet.2005.01.001

Behavioral questions were more reliable than questions with emotional content. Measurements among younger respondents were also more reliable.

JEL classification: C00; C42; C80

Keywords: Egocentered networks; Measurement; Reliability; Validity; Social support

1. Introduction

Social network analysis has become very important in many research fields (e.g., sociology, political science, economics, anthropology, organizational sciences, and medicine). One of the most important fields where social network analysis is used is social support, where it has been reported to influence physical and mental health, as well as general well-being (e.g., Hirsch, 1981; House, 1981; Gottlieb, 1983; Berkman, 1985; Wenger, 1994; Latkin et al., 1995; Samuelsson, 1997).

Studying the measurement quality of social network measurement instruments is important because it has the potential to provide information relating to the factors that influence the reproducibility of results and the validity of the underlying concepts. This is perhaps even more important for egocentered network data, since data about the network and its characteristics and the characteristics of network members are usually given by the respondent (ego).

In substantive research on social networks, different network characteristics (such as network size, structure and composition) and characteristics of network members (such as gender, closeness, importance, frequency of contact) are studied. Within this paper, the aim is to estimate the reliability and validity of frequently used name interpreters. Further, since data about the characteristics of ties are important explanatory variables in social support research and are, moreover, usually reported by the ego, it is very important to know the extent to which the measures are reliable and valid. However, as the intended unit of analysis is egocentered network as a whole and not individual ego-alter ties, the variables are defined as averages of name interpreters for each egocentered network. The use of averages is further justified by the fact that averages of these variables are often used in the substantive research on social support. Therefore, the reliability and validity of the averages for these variables were studied.

2. Quality of measurement

The main purpose of scientific research is the discovery of laws on the basis of which interpretation and prediction of phenomena are possible. In this endeavor the quality of the measurement instruments (their reliability and validity) with which we obtain empirical data for the attainment of this purpose, is of crucial importance. In general, reliability of data can be defined as the ability to obtain the same (or at least very similar) scores at repeated measurements on the same units, on the assumption that the true scores have remained the

same. Validity, on the other hand, gives the assurance that we have really measured the concept that we intended to measure.

In the social sciences (as well as in the field of social network analysis) the most frequently used measurement instrument is a survey. Research into the quality of survey data about attitudes has a long tradition in social science methodology (e.g., Cantril, 1944; Payne, 1951; Sudman and Bradburn, 1982; Hippler et al., 1987; Converse and Presser, 1988; Schwarz and Sudman, 1992, 1994, 1996; Sudman et al., 1996). The question of the quality of social network data was first systematically dealt with in the 1970s (Killworth and Bernard, 1976, 1979/1980; Bernard and Killworth, 1977; Bernard et al., 1979/1980, 1982, 1985). The main finding of these studies was that people are generally very inaccurate in reporting on their past interactions with other people. Later studies (e.g., Freeman and Romney, 1987; Freeman et al., 1987; Corman and Bradford, 1993) confirmed this finding, but also showed that, on the other hand, people remember long-term or typical patterns of interaction with other people rather well. In addition, it has been shown that the accuracy of reporting about interactions is also influenced by the frequency of interaction and by the reliability of an individual respondent. Respondents who were in more frequent contact with other group members had more accurate reports about behavior within the group, and respondents with higher reliability were also shown to report actual interactions more accurately (Romney and Faust, 1982; Romney and Weller, 1984).

Several other studies of the quality of measurement of egocentered and complete networks have also shown that richer data can be obtained with the free recall method. On the other hand, the fixed choice method can be shown to provide accurate information on the most important relationships (Hammer, 1984; Hlebec, 1993). In a comprehensive literature review, Brewer (2000) has found that forgetting to name people in recall-based elicitation of social (complete) networks is a potentially significant problem when collecting such data. People seem to be more likely to forget weak ties. On the other hand, people may also forget a significant proportion of close ties.

In many systematic and comprehensive studies of measurement quality (e.g., Ferligoj and Hlebec, 1995, 1998, 1999; Hlebec and Ferligoj, 1996, 2001; Hlebec, 1999) it has also been found that the quality of measurement (especially reliability) of complete networks is influenced by the dimension of social support, method order, time between repeated measurements, type of measurement scale, mood and the degree of change in the mood.

After the list of alters is obtained with name generators, name interpreter questions can be asked in two ways. One way (“by alters”) is to take each alter individually and to ask all questions about him/her, going alter by alter until the end of the list of alters. The other way (“by questions”) is to take the question and ask this question for all alters on the list, going question by question until the end of the list. Recent studies of reliability and validity of measurement in egocentered social network data have shown that it is advisable to measure tie characteristics (e.g., frequency of contact, closeness, importance of network and members, frequency of negative interactions) within a telephone interview combined with data collection by alters. Further, telephone interviews produce: (a) high reliability and good validity; (b) more named networks than personal interviews, and (c) considerable cost savings over personal interviews (Kogovšek et al., 2002).

3. Hypotheses

It is hypothesised that the characteristics of ties are measured more reliably by alters than by the question data collection technique (H1a). It is expected that cognitively more demanding questions (e.g., frequency of contact between ego and his/her alters) are more prone to measurement errors in telephone than in personal interviews. On the other hand, with the lack of the physical presence of the interviewer, telephone interviews may be slightly more anonymous than personal interviews, which could produce more socially desirable responses to sensitive questions (e.g., feelings of closeness, frequency of alters upsetting the ego) (e.g., Hippler et al., 1987). Since the measurement method is a combination of data collection mode and data collection technique, we expect that the combination personal interview/by alters would yield the highest reliability and validity of measurement for cognitively more demanding questions (H1b). On the other hand, since telephone interviews are more anonymous, we expect sensitive questions to be most reliably and validly measured by telephone/by alters (H1c).[1]

One characteristic that may also affect the quality of measurement of the characteristics of ties is network size. With larger networks the respondent's effort increases correspondingly; therefore it can be expected that the characteristics of ties will be more reliably measured in smaller networks. On the other hand, a larger named network usually consists of a larger proportion of less important, weaker, more distant ties. In comparison to strong ties, the respondent usually does not know weak ties as well; they are not as important and close; therefore measurement method effects may be more prominent, which in turn lowers validity. Therefore, we expect that the characteristics of ties will be more validly measured in smaller networks (H2).

The quality of measurement of egocentered networks can also be affected by respondents' personal characteristics. Research shows (e.g., Fischer, 1982; Campbell et al., 1986; Marsden, 1987; van der Poel, 1993) that networks can differ with reference to different personal characteristics such as gender, age, education, race/ethnicity. Network size usually decreases with age (Marsden, 1987; van der Poel, 1993) and increases with education (Fischer, 1982; Campbell et al., 1986; Marsden, 1987; van der Poel, 1993). Research also shows (e.g., Wellman and Wortley, 1990) that women, to a greater degree than men, provide emotional support and social companionship and that they are "traditionally" found in the role of someone who takes care of personal relationships.

Several studies show (e.g., Turner, 1994) that women have more frequent contacts with their network members and have a tendency for these relationships to be more emotionally intimate. Women more than men name people that they know very well or to whom they feel very close (i.e., they have a greater proportion of such ties in their networks) (Fischer, 1982; Hammer, 1984; Antonucci et al., 1998). Therefore, it is to be expected that the characteristics of ties in women's networks will be both more reliable and valid than the tie characteristics in men's networks (H3a). On the other hand, some research also shows (an overview in Antonucci, 1985) that women tend to have more heterogeneous networks than men, which could, in turn, cause lower reliability and validity of measurement of women's networks and/or their characteristics (Table 1).

[1] For a detailed theoretical argumentation see Kogovšek et al. (2002: 3–5).

Table 1
The list of hypotheses

Hypothesis	Predictors
Measurement instrument	
H1a	Higher quality of measurement with by alters than by questions data collection technique
H1b	Higher quality of measurement of cognitively demanding questions with the personal interview/by alters
H1c	Higher quality of measurement of sensitive questions with the telephone interview/by alters
Network size	
H2	Higher quality of measurement in small networks
Personal characteristics	
H3a	Higher quality of measurement among female respondents
H3b	Higher quality of measurement among younger respondents
H3c	Higher quality of measurement among more educated respondents
Personality characteristics	
H4a	Higher quality of measurement among extraverts
H4b	Higher quality of measurement among emotionally stable respondents

Responding to questions about network members and their characteristics is cognitively a very demanding task for the respondent—one which requires processing a large amount of information (different tie and alter characteristics) and relatively complex, difficult and potentially time consuming cognitive operations (assessing "average" quality of support, closeness, duration of the relationship and so on). In this context we expect that the reliability and validity of measurement of tie characteristics will be higher for younger respondents (H3b). For the same reasons, we also expect that the reliability and validity of measurement will be higher for more educated respondents (H3c).

It has already been established (Hlebec, 1999; Hlebec and Ferligoj, 2001) that current mood affects the stability of measurement in complete networks. Respondents who are in a positive mood name a larger number of alters than respondents who are in a negative mood. The measured network is also affected by a change in mood between two measurements and it is also known that differences in personality characteristics affect the perception of social support (e.g., Procidano and Heller, 1983; Sarason et al., 1983; Sarason and Sarason, 1985; Vinokur et al., 1987; Newcomb and Keefe, 1997; Hooker et al., 1998; Mongrain, 1998; Nolen-Hoeksema and Christopher, 1999). Some research also shows that some personality characteristics (e.g., extraversion, neuroticism (emotional instability), depression and anxiety) affect the capacity for information processing (an overview in Matthews and Wells, 1999; McLaughlin and Eysenck, in Hall et al., 1998). With depressed respondents, negative emotions require a greater share of the cognitive capabilities; therefore only a smaller part is available for the processing of other cognitive operations than is the case with non-depressed respondents, especially with operations that demand a high level of focus (concentration) and/or are cognitively very demanding (both of which hold for the survey situation). This decrease in cognitive functioning is seen as related to the self focusing of attention" (Ingram, 1990 in Matthews and Wells, 1999). The intensive self-focus of attention causes performance deficits that are related to negative emotions.

From the previous research it is evident that there are complex connections among perceptions of psycho-social factors (such as social support, current mood, stress, depression and personality characteristics) and information processing mode and the related quality of measurement. Psycho-social characteristics appeared to be an important factor in most of those studies; hence, it is expected that they will also affect the quality of measurement in social support networks. Casciaro (1998), in a study of the effect of extraversion on the accuracy of reporting about friends' use of advice within a work environment, found a weak positive correlation between extraversion and the accuracy of reported friendship networks. As a possible reason she stated that this correlation is more likely in the context of more personal relationships than in the presence of more instrumentally oriented relationships, e.g., advice related to work. Relationships in which an individual seeks social support are usually more personal and intimate; therefore we expect that personality characteristics, such as extraversion and emotional stability, will have an effect on the reliability and validity of the measurement of these relationships. So therefore, it is expected that the reliability and validity of measured tie characteristics will be higher among extraverts than for introverts (H4a). Similarly, the quality of measurement is expected to be higher for more emotionally stable individuals than for the emotionally unstable (H4b).

4. MTMM models

We approach the problem of estimating data quality from the standpoint of the well-known and widely used Campbell–Fiske MTMM approach (Cambell and Fiske, 1959). Since the introduction of path analytic models within SEM, MTMM matrices are usually analyzed by Confirmatory Factor Analysis (e.g., Bollen, 1989).

A number of MTMM models have been formulated and tested (e.g., Althauser et al., 1971; Alwin, 1974; Werts and Linn, 1970; Browne, 1984, 1985; Marsh, 1989; Saris and van Meurs, 1990; Saris and Andrews, 1991; Ferligoj et al., 1995; Saris and Münnich, 1995; Scherpenzeel, 1995c; Coenders and Saris, 2000).

The MTMM formulation that appears to be the most useful, at least in the present context, is the true score model as proposed by Saris and Andrews (1991), which has already been succesfully used on social network data (Ferligoj and Hlebec, 1999; Kogovšek et al., 2002).[2] The true score model is defined as follows:

$$Y_{jk} = h_{jk} T_{jk} + e_{jk} \tag{1}$$

$$T_{jk} = v_{jk} F_j + m_{jk} M_k \tag{2}$$

where Y_{jk} is the measured variable (trait F_j measured by the kth method), T_{jk} is the stable component of the response Y_{jk} (the "true score"), F_j is the trait and M_k is the variation in scores due to the kth method (Fig. 1).

[2] For a detailed description of the model, see the references mentioned.

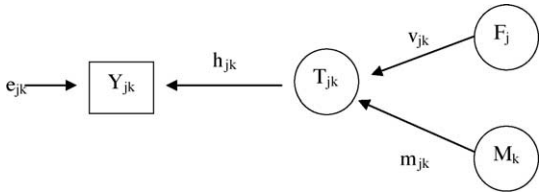

Fig. 1. True score measurement model.

If all the variables are standardized, the standardized parameters represent the following:

- h_{jk} is the reliability coefficient (h^2_{jk} being the test-retest reliability),
- v_{jk} is the validity coefficient (v^2_{jk} representing the validity of the measure); and
- m_{jk} is the method effect where $m^2_{jk} = 1 - v^2_{jk}$, which means that the method effect is equal to the invalidity of the measure.

At this point it should be stressed that validity within this model is defined as the absence of method effect and could be named empirical validity. Therefore, within this model we can strictly speak only of validity (and reliability) coefficients, which is a much more limited concept of validity than it is otherwise often used (e.g., construct validity). Reliability is defined as the absence of random measurement error, which is consistent with the classical test theory (Lord and Novick, 1968).

5. The design of the study

The standard true score model requires answering the selected questions at least three times. This is a tedious task for respondents. Therefore, we decided to use a form of split ballot MTMM design, first proposed by Saris (1999). In his design, respondents were randomly assigned into two groups with different combinations of methods, but each group used only two methods. In the first measurement, all respondents received the first method, but in the second measurement, one group received the second and the other group the third method. In our study, a design similar to Saris' (1999) was used, but with three groups,[3] each with two out of the three methods, an approach which is displayed in Table 2. The methods used were a combination of the data collection mode (face-to-face and telephone) and data collection technique (by alters and by questions).[4]

6. Data

The data were collected between March and June 2000 by computer-assisted telephone interview (CATI) and computer-assisted personal interview (CAPI) for a representative

[3] For advantages of the three-group design over the two-group design, see Kogovšek et al. (2002).

[4] We are aware of the limitations of the design of the study. Some suggestions for improvements were presented in Kogovšek et al. (2002) and were actually tested by Saris et al. (2001).

Table 2
The design of the study

Group	*N*	First interview	Second interview
1	320	Face-to-face/by alters	Telephone/by alters
2	311	Face-to-face/by alters	Telephone/by questions
3	402	Telephone/by alters	Telephone/by questions

sample of 1033 inhabitants of the city of Ljubljana, Slovenia. These respondents produced 7223 alters. The time span between the two measurements was one week.[5]

The following name generators were used:

1. From time to time people borrow something from other people, for instance, a piece of equipment, or ask for help with small jobs in or around the house. Who are the people you usually ask for this kind of help? (material support).
2. From time to time people ask other people for advice when a major change occurs in their life, for instance, changing jobs or a rather serious accident. Who are the people you usually ask for advice when such a major change occurs in your life? (informational support).
3. From time to time people socialize with other people, for instance, they visit each other, go together on a trip or to a dinner. Who are the people with whom you usually do these things? (social companionship).
4. From time to time people discuss important personal matters with other people, for instance: (a) if they quarrel with someone close to them (b) when they have problems at their work or something similar. Who are the people with whom you discuss personal matters that are important to you? (emotional support).
5. Suppose you found yourself in a situation where you needed a large sum of money, for instance, five average monthly wages (approximately 500.000 tolars). Whom would you ask to lend you the money (a person, not an institution, e.g., a bank)? (financial support).

The traits used in this study are represented by two widely used measures of tie strength (for an overview of this topic, see Marsden and Campbell, 1984): the frequency of contact of the ego with each alter and the ego's feelings of closeness towards each alter. The third trait is represented by a measure of the negative aspects of social relationships: the frequency of the alter upsetting the ego.[6]

However, as the units of analysis here are egocentered networks as a whole and not individual ego-alter ties, the traits are defined as averages of these three variables for each egocentered network. So the traits in this study actually represent the average frequency of contact of an ego with his/her alters, the average closeness between the ego and his/her alters and the average frequency of alters upsetting the ego within these networks.

Extraversion and emotional stability scales were used for measuring the respondents' personality characteristics. Since their initial theoretical definition and empirical operationalization (Eysenck, 1947; Eysenck and Eysenck, 1968), these two personality dimen-

[5] As the measured networks are usually not fixed and stable, but are rather dynamic, some possible real changes could happen between the measurements besides measurement error and be reflected in the results.

[6] Question wording, see Kogovšek et al. (2002).

sions have been consistently confirmed in many different studies and cultural settings. In our study, the 10-item versions of the IPIP scales (Goldberg *International Personality Item Pool* (http://ipip.ori.org/ipip)) were used.[7]

7. Meta analysis of the factors affecting quality of measurement of egocentered networks

Meta analysis is a secondary analysis of previously published results, which are used to compare, connect and summarize findings from many studies within a given research field (e.g., Glass in Scherpenzeel, 1995c). With a meta analysis, we can also analyze the effect of several factors on the results within a single study.

A special case of meta analysis involves meta analyses of MTMM studies with the aim of exploring the factors affecting the quality of measurement. Within MTMM meta analyses the method most frequently used was multiple classification analysis (Andrews et al., 1973). This method was successfully employed in a number of meta analyses analyzing the effects of survey measurement instruments' methodological properties (such as the type of scale, the length of questionnaire, the sensitivity of the topic, data collection mode and so forth) on the quality of attitude and opinion measures (e.g., Ferligoj et al., 1995; Scherpenzeel, 1995a, 1995b, 1995c; Költringer, 1995), as well as network data (Ferligoj and Hlebec, 1999; Hlebec, 1999; Hlebec and Ferligoj, 2002).[8]

Multiple classification analysis (Andrews et al., 1973) is a multivariate method, by which relationships between multiple independent variables (or predictors) and a dependent variable are analyzed. Multiple classification analysis gives us the following information:

- The overall mean and group means of the dependent variable for each combination of categories of predictors.
- Tests of significance of the effects of single predictors as well as of interactions between them.
- The effect of each predictor is shown by parameter β, which tells us the effect of the predictor if other predictors are held constant; thus the rank order of βs shows us the relative importance of a single predictor in explaining the dependent variable.
- Deviations from the total mean of the dependent variable for each category of a predictor (therefore, how much would the total mean of the dependent variable increase or decrease as a result of the effect of a certain predictor).
- The percentage of explained variance for all predictors included in the analysis (R^2).

[7] Both scales have fairly good measurement characteristics. On American data the Cronbach α was .87 for extraversion and .86 for emotional stability. In this study, the Cronbach α was .77 for extraversion and .82 for emotional stability. Factor analysis with principal axis factoring method and oblimin rotation on our data showed a clear 2-factor solution with the exception of one item (*Am relaxed most of the time*), which should have measured emotional stability, but had relatively high weights on both factors; therefore this item was excluded from further analyses. All the items used can be seen in Appendix A.

[8] More about MTMM meta analyses in Saris and Münnich (1995), Scherpenzeel (1995c), Ferligoj et al. (1995, pp. 142–152) and Hlebec (1999, pp. 59–69).

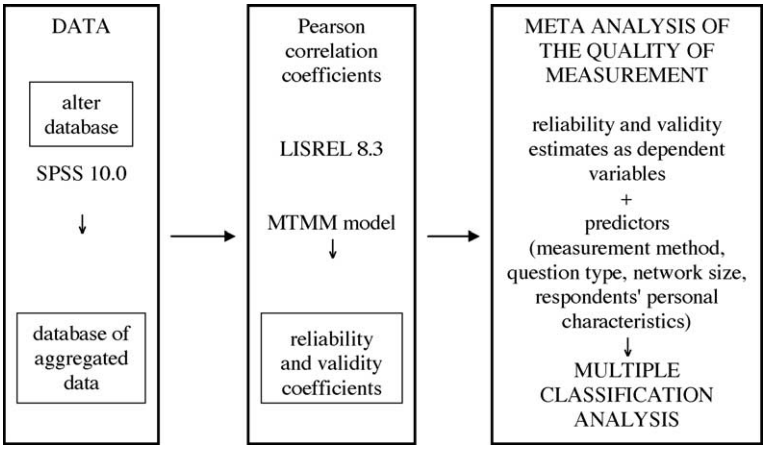

Fig. 2. The procedure of the meta analysis.

In the meta analysis, the input data (units of analysis) are not the respondents' responses, as in usual analyses, but the measurement instrument—a specific combination of the measurement instrument characteristics (e.g., measurement method, question type, respondent characteristics) together with the reliability and validity coefficient for the particular measurement instrument. Thus, in the meta analysis, reliability and validity coefficients are dependent variables, and measurement instrument characteristics are independent variables (predictors). An example of a data matrix for a meta analysis can be seen in Appendix B.

The procedure for the analysis is shown in Fig. 2. Firstly, the data were aggregated on the level of egocentered networks. In the analysis of egocentered networks the unit of analysis is the network (not an individual respondent or individual ego-alter tie); hence, researchers often interpret aggregated data, e.g., averages. Therefore, in the current study the variables used were the averages from egocentered networks (average frequency of contact, average closeness between ego and his/her alters and average frequency of alters upsetting the ego). Then the respondents were divided into subsamples. In the next step, correlation coefficients between the aggregated variables, each measured by each of the three methods, were calculated for each subsample (MTMM matrix). On the basis of the MTMM matrices, reliability and validity coefficients were obtained using LISREL 8.3. In the last step, reliability and validity coefficients were used as dependent variables within the multiple classification analysis. The predictors used were network size, measurement method, characteristics of the questionnaire and respondent characteristics (gender, age, education, and personality traits). For instance, in the meta analysis, represented by the data matrix in Appendix B, respondents were divided into four subsamples by gender and age. Considering also that each of the three aggregated variables is measured by each of the three measurement methods, we get 36 units of analysis in this particular meta analysis.

When doing the meta analysis of the factors affecting the quality of measurement, it would be ideal to include all relevant factors together. But this ideal is circumscribed by

at least three limitations. One problem that can arise is that some predictors may be a linear combination of other predictors, or at least correlate with each other very strongly.[9] There has to be a sufficient number of cases (in this case reliability and validity estimates) in comparison with the number of predictors (Andrews et al., 1973; Scherpenzeel, 1995c).[10] The first limitation does not hold in our case, but the second and third do—a relatively low number of cases and the sample size. For instance, if we want to test the effects of respondent characteristics (e.g., gender, age, education, and personality traits), we have to form subsamples for all possible combinations of those characteristics. On the other hand, those subsamples have to be large enough to obtain stable reliability and validity estimates.[11] In our case, the split ballot research design has another limitation. For the calculation of reliability and validity estimates we needed three correlation matrices, one for each group of respondents; therefore each of the subsamples for the meta analysis is further divided into three groups. This means that with a total sample size of about 1000 cases, only two predictors besides measurement method can be included in a meta analysis. This was the reason for conducting several meta analyses with different combinations of predictors in order to study the effects and their interactions as thoroughly as possible.[12]

In the analysis interactions between predictors can also occur. When statistically significant interactions occurred in our analyses,[13] they were examined according to the suggestion by Andrews et al. (1973) that predictors in interaction be merged into a single predictor.

In the next two Subsections the results of the meta analyses are presented. All the meta analyses are pooled into Tables 3 and 4 , the first showing the results for the reliability and the latter for the validity of measurement. For each meta analysis the average reliability/validity coefficient and its standard deviation is shown as well as the proportion of explained variance of all included predictors. For each included predictor the strength of its effect (β), the deviation from the average reliability/validity coefficient for each category of the predictor (Dev.) and reliability/validity coefficients (Rel./Val. Coeff.)[14] are shown.

7.1. Meta analyses of the reliability of measurement

In this section the results of the meta analyses of the reliability of measurement are presented (Table 3). In general, it can be seen that the effects of the included predictors are not statistically significant. Only the effects of the measurement method (in two meta analyses),

[9] More about this problem in Scherpenzeel (1995c, pp. 32–34).

[10] There should be many more cases than the degrees of freedom of the model. Degrees of freedom can be obtained by subtracting the number of predictors from the sum of all categories of all predictors (Andrews et al., 1973).

[11] There is no final criterion for the sample size for such models. It is advisable to have several hundred cases (Ferligoj et al., 1995), but it also seems that, if this is not possible, about 100 cases are sufficient.

[12] The combination of predictors in each meta analysis was not chosen at random, but on the basis of well-thought theoretical considerations (e.g., which predictors could be expected to be in interaction; the fact that different combinations of predictors may produce different results).

[13] Owing to sample size limitations, only interactions of two predictors could be examined.

[14] These are obtained by subtraction of the deviation from the average reliability/validity coefficient.

Table 3
Meta analyses of the reliability of measurement

	Meta analysis 1 $\bar{X}_{rel} = .832$, $\sigma = .071$, $R^2 = .357$			Meta analysis 2 $\bar{X}_{rel} = .828$, $\sigma = .089$, $R^2 = .239$			Meta analysis 3 $\bar{X}_{rel} = .834$, $\sigma = .067$, $R^2 = .178$		
	β	Deviation	Reliability coefficient	β	Deviation	Reliability coefficient	β	Deviation	Reliability coefficient
Method	.446*			.327			.416		
Personal/by alters		.011	.843		.009	.837		.007	.841
Telephone/by alters		.031	.863		.030	.831		.029	.863
Telephone/by quest.		−.042	.790		−.039	.789		−.036	.798
Network size[a]	.240								
1–5		−.017	.815						
6+		.017	.849						
Type of question	.317*								
Behavior		.031	.863						
Emotional		−.016	.816						
Age				.292*					
40 years or less					.026	.854			
41+					−.026	.802			
Gender				.216					
Male					−.019	.809			
Female					.019	.847			
Education							.068		
Up to compl. second.								−.004	.830
College or more								.004	.838
Extraversion									
Introverted									
Extraverted									
Emotional stability									
Emotionally unstable									
Emotionally stable									

Table 3(*Continued*)

	Meta analysis 4 $\bar{X}_{rel}$ = .832, σ = .095, R^2 = .158			Meta analysis 5 $\bar{X}_{rel}$ = .829, σ = .091, R^2 = .118			Meta analysis 6 $\bar{X}_{rel}$ = .826, σ = .086, R^2 = .164		
	β	Deviation	Reliability coefficient	β	Deviation	Reliability coefficient	β	Deviation	Reliability coefficient
Method	.332			.290			.399*		
Personal/by alters		.016	.848		.006	.835		.007	.833
Telephone/by alters		.028	.860		.029	.858		.037	.863
Telephone/by quest.		−.043	.789		−.034	.795		−.044	.782
Network size[a]	.036			.052					
1–5		−.003	.829		−.005	.824			
6+		.003	.835		.005	.834			
Type of question									
Behavior									
Emotional									
Age	.214								
40 years or less		.020	.852						
41+		−.020	.812						
Gender				.176					
Male					−.016	.813			
Female					.016	.845			
Education									
Up to compl. second.									
College or more									
Extraversion							.066		
Introverted								−.006	.820
Extraverted								.006	.832
Emotional stability							.020		
Emotionally unstable								.002	.828
Emotionally stable								−.002	.824

[a] Since there were no respondents with network sizes of less than six in meta analyses four and five, we decided to raise the boundary between a small and a large network by one. Therefore, a network consisting of one to six alters is considered small and a network consisting of seven or more alters is large.

* $.05 < p < .10$.

Table 4
Meta analyses of the validity of measurement

	Meta analysis 1 $\bar{X}_{val} = .963$, $\sigma = .026$, $R^2 = .426$			Meta analysis 2 $\bar{X}_{val} = .961$, $\sigma = .038$, $R^2 = .562$			Meta analysis 3 $\bar{X}_{val} = .958$, $\sigma = .025$, $R^2 = .571$		
	β	Deviation	Validity coefficient	β	Deviation	Validity coefficient	β	Deviation	Validity coefficient
Method	.358***			.601***			.721***		
Personal/by alters		−.013	.950		−.031	.930		−.024	.934
Telephone/by alters		.007	.970		.014	.975		.007	.965
Telephone/by quest.		.006	.969		.017	.978		.017	.975
Network size[a]	.460***								
1–5		.012	.975						
6+		−.012	.951						
Type of question	.295***								
Behavior		.011	.974						
Emotional		−.005	.958						
Age				.192**					
40 years or less					.007	.968			
41+					−.007	.954			
Gender				.405***					
Male					.015	.976			
Female					−.015	.946			
Education							.226		
Up to compl. second.								−.006	.952
College or more								.006	.964
Extraversion									
Introverted									
Extraverted									
Emotional stability									
Emotionally unstable									
Emotionally stable									

Table 4(*Continued*)

	Meta analysis 4 $\bar{X}_{val} = .954$, $\sigma = .035$, $R^2 = .464$			Meta analysis 5 $\bar{X}_{val} = .955$, $\sigma = .036$, $R^2 = .401$			Meta analysis 6 $\bar{X}_{val} = .956$, $\sigma = .035$, $R^2 = .449$		
	β	Deviation	Validity coefficient	β	Deviation	Validity coefficient	β	Deviation	Validity coefficient
Method	.529***			.516***			.503***		
Personal/by alters		−.025	.929		−.023	.932		−.023	.933
Telephone/by alters		.008	.962		.002	.957		.019	.975
Telephone/by quest.		.018	.972		.021	.976		.004	.960
Network size[a]	.307***			.296**					
1–5		.011	.965		.011	.966			
6+		−.011	.943		−.011	.944			
Type of question									
Behavior									
Emotional									
Age	.301***								
40 years or less		.010	.964						
41+		−.010	.944						
Gender				.218**					
Male					.008	.963			
Female					−.008	.947			
Education									
Up to compl. second.									
College or more									
Extraversion							.399***		
Introverted								−.014	.942
Extraverted								.014	.970
Emotional stability							.191**		
Emotionally unstable								−.007	.949
Emotionally stable								.007	.963

[a] Since there were no respondents with network sizes of less than six in meta analyses four and five, we decided to raise the boundary between a small and a large network by one. Therefore, a network consisting of one to six alters is considered small and a network consisting of seven or more alters is large.

** $.01 < p < .05$.

*** $p < .01$.

the type of question and respondent age on reliability proved to be marginally statistically significant. There is a tendency for measurement to be most reliable for telephone/by alters and least reliable for telephone by questions. Both methods that include data collection by alters seem to increase reliability, while the method including data collection by questions lowers it. Therefore, we get some confirmation of hypothesis 1a—that data collection is more reliable by alters than by questions. We can also see that, regardless of which predictors are included in the meta analysis, the effect of the measurement method on reliability is the greatest of all.

If the respondent is asked about behavior, a slightly more reliable measurement can be expected. If s(he) is asked about matters with emotional content, the average reliability tends to be a little lower.

The effect of the network size[15] on reliability is non-significant. But there is a tendency for the characteristics of ties to be measured more reliably in larger networks. This result contradicts hypothesis 2. One possible explanation could be the interference of another variable. There is a slightly larger proportion of older respondents in the sample, and these tend to have smaller networks. Therefore, reliability in smaller networks may be due to the hidden effect of age on reliability, which could "artificially" lower the reliability for smaller networks. However, there was no interaction between network size and age. Another reason may lie in the decision to use average network size as the cut point between small and large networks. It may be that this limit is not the most suitable considering cognitive information processing abilities, and in any event the "jump" in reliability may be happening at a different limit in network size.[16] It is also possible that measurement errors become more prominent in smaller networks than in larger ones, e.g., if the respondent leaves out one or two alters in one of the measurements, and hence leading to an increase in measurement error within smaller networks. It is therefore, advisable to study ways to minimize this possible effect.

For the reliability of measurement there was also a significant interaction between the measurement method and the type of question. Additional analysis of this interaction shows that a question about behavior is measured more reliably in face-to-face interviews and less reliably in both versions of the telephone interview. Both questions with emotional content are measured more reliably within the telephone by alters' condition. These results confirm hypotheses 1b and c—that cognitively more demanding questions would be measured more reliably by face-to-face mode, and that questions that are potentially more sensitive would be measured more reliably in the telephone by alters condition. The explanation may be that the latter data collection mode is more anonymous, and, in combination with data

[15] Respondents were divided into two groups by average network size, which was between five and six alters. Therefore a network with one to five alters was considered small, and a network with 6 or more alters was considered large.

[16] An additional analysis was done with only network size and measurement method as predictors. According to the distribution of network size, respondents were divided into three groups: those with small networks (one to four alters), those with middle sized networks (five to six alters) and those with large networks (7+ alters). The results were similar. The effects on reliability of measurement were statistically non-significant, but there was a tendency for reliability to be smaller in small networks and higher in middle sized and large networks. Therefore it seems that a wrongly placed limit between small and large networks is not the reason for the contradictory results in the case of reliability of measurement.

collection by alters is cognitively less demanding since there are fewer context effects. Certainly these interactions should be studied more thoroughly in the future with a larger number of questions of both types.

There was a marginally statistically significant effect of age on reliability of measurement.[17] Consistent with hypothesis 3b, lower reliability can be expected with older respondents and higher reliability with younger ones.

7.2. Meta analyses of the validity of measurement

In general, we can see that there are much stronger effects, which are all also statistically significant in the case of the validity of measurement (Table 4). Measurement method seems to have the strongest effect on the validity coefficients (except in the case of meta analysis 1, where the effect of the network size is stronger) and it is consistent in all meta analyses (see deviations from the average validity). If the telephone method is used, an increase in validity can be expected. If the face-to-face method is used, validity will decrease, even if it is combined with a cognitively less demanding data collection by alters. This result shows that the greater anonymity of the telephone mode may indeed contribute to an increase in the validity of measurement of characteristics of ties in egocentered networks.

If a respondent has a smaller network, it can be expected that average validity will increase by .011–.012 ($\bar{X}_{\text{val}}$ = .974–.975), and if s(he) has a larger network, average validity will decrease by the same amount. This result confirms hypothesis 2—that the characteristics of ties are measured more validly in smaller networks. The reason for the lower validity in larger networks may lie in the fact that larger networks usually contain a larger proportion of weak ties.[18] It could be argued that the ego is not so familiar with weaker ties and consequently this may lower the validity of measurement within the whole network; since measurement error may be more pronounced.

The type of question also has significant effect on the validity of measurement. If we ask about frequency of contact a slightly higher validity is shown and, for questions with emotional content, slightly lower validity can be expected.

Contrary to hypothesis 3a, it turns out that women had less validly measured characteristics of ties than men. A possible reason could lie in the fact that women have larger networks than men do, probably containing a larger proportion of weak ties and this might enhance the effect of the measurement method and thus lower the validity.

Validity was lower with older respondents, a finding that confirms hypothesis 3b. The reasons for such a result may be two-fold. First, older respondents may give less valid responses owing to factors such as memory or hearing problems. Second, older respondents often have more weak ties in their networks, e.g., extended family kin,

[17] Respondents were divided into two groups by average age in the sample (40 years).

[18] That this is indeed the case can be observed from the Pearson correlation coefficients between network size and average feelings of closeness ($r = -.22$ in the first, and $r = -.16$ the second measurement), and between network size and average importance of alters ($r = -.21$ in first, and $r = -.13$ in the second measurement). Thus, the larger the network, the less close an ego feels toward the alters and the less they are important to her/him.

neighbors, professional helpers, many of whom they may not know well. There may also be an interaction between gender, age and network size. There is some evidence (e.g., Fischer, 1982) that older women may be socially more active and therefore have larger networks than men, and that older men tend to be among the most socially isolated social groups.

Also both personality traits[19] had a statistically significant effect on the validity estimates. Introverts have a lower validity of measurement, and the same was true for emotionally unstable respondents, therefore confirming hypotheses 4a and 4b. Possible explanations for this result may be that these respondents have less stable relationships with other people, or that their perception of these relationships is less stable. If we take a look at the effect of both personality characteristics, it can be seen that the average validity is slightly higher for extraverts who were emotionally stable (.977) vis-à-vis the extraverts who were emotionally unstable (.963). On the other hand, introverts who were emotionally stable had a slightly higher average validity (.949) vis-à-vis emotionally unstable introverts (.935). These results are in line with the results of the McLaughlin and Eysenck experiment (in Hall et al., 1998), which showed that when performing demanding tasks, emotionally stable extraverts obtain the best results and emotionally unstable introverts the worst, with the other two groups in between.

In the case of validity of measurement, there is a significant interaction between measurement method and network size. Additional analysis shows that characteristics of ties in smaller networks are measured with a somewhat higher degree of validity by face-to-face and telephone mode combined with data collection by alters. It seems that with smaller networks the data collection technique (by alters/by questions) is more important than the data collection mode (telephone/face-to-face). Since smaller networks contain important alters that the respondent knows very well, it can therefore be expected that s(he) would give consistent answers about the characteristics of ties, regardless of the measurement method. But it is quite possible that context effects are stronger when the data is collected by questions, since the respondent answers the same question for all alters first and not vice versa. For instance, if we consider a question relating to feelings of closeness, it is more likely that the respondent would compare the alters while answering the question. In that case, it would not really be the actual feelings of closeness towards each individual alter that would be measured, but the feelings of closeness relative to previous alters on the list.

In larger networks the characteristics of ties are measured with the least validity in the face-to-face mode and with more validity by telephone mode, regardless of the data collection technique used. It seems that with larger networks the aforementioned context effects are no longer so prominent.

There are also significant interactions between age and gender and gender and measurement method. For instance, additional analysis in the case of meta analysis 2 showed that the difference in average validity between older and younger men (.972 and .979) is smaller than the difference between younger and older women (.963 and .928). Similarly,

[19] The respondents were divided into introverted and extraverted by standardized factor scores on the extraversion factor. Respondents with negative factor scores were designated as introverted, and respondents with factor scores of 0 or positive, as extraverted. The same was done for emotional stability.

the difference in average validity between the face-to-face method (.928), on the one hand, and telephone methods (.992) on the other, was greater in the male group than in the female group (average validity estimates ranging from .930 to .957). There was no great difference between young males and young females. There was also not a large difference between younger and older men. However, considerable difference appears if we compare younger and older women and older men and older women. The average validity is lower for older women and higher in other categories.

Additional analysis showed that responses from females had a slightly higher validity in the face-to-face mode, whereas men have higher validity for the telephone mode. The differences between both methods (alters and questions) in the telephone mode (.999) and the face-to-face mode (.928) are larger for men than women (average validity coefficients between .930 and .957). It seems that with men, the difference between telephone and face-to-face interviewing becomes more prominent. Being interviewed about personal relationships by telephone suits both genders better, but face-to-face interviews suit women better than men.

Two interactions regarding personality traits were statistically significant: (1) emotional stability and measurement method and (2) extraversion, emotional stability and measurement method. Only the first was examined because of the limited sample size. Emotionally unstable respondents had on average slightly higher validity values when data was collected by alters, and on average lower validity scores when the data was collected by questions. On the other hand, emotionally stable respondents had on average lower validities in both the face-to-face interviews and in the telephone condition. We have already seen that the data collection method by questions seems to be cognitively more demanding, and that context effects may be stronger, a situation to which emotionally unstable respondents tend to be more sensitive. On the one hand, emotionally unstable respondents may have a high level of arousal due to the potentially demanding nature of the task and, as a result, have an overly high level of motivation. Further, they may be more sensitive to the context of the current interviewing situation, owing to other characteristics of their personality such as an inclination to anxiety, mood swings and impulsive emotional reactions, etc.

8. Conclusion

In this study a number of factors that may affect the quality of measurement of average characteristics of ties in egocentered networks were tested within a series of meta analysis. It should be stressed again that validity within the MTMM model used means consistency across measurement methods and it is therefore a more limited concept than validity in the general sense.

It can be seen that the effects of all those factors were consistent. The telephone/by alters measurement method appeared to be the most reliable and the telephone/by questions measurement method the least reliable. The result provided some support for the hypothesis that data collection by alters is more reliable than data collection by questions.

Both telephone methods were consistently more valid than personal interviews (in all cases the effects were statistically significant). Therefore, we have some support for the

hypothesis that measurement by telephone is more valid than that by personal interview. The reason for this may lie in the relative sensitivity of the topic and the relative anonymity of the telephone method.

The study also showed that the data collection technique (by alters/by questions) mostly affected the reliability of measurement, whereas the data collection mode appeared mostly to affect the validity of measurement. The telephone/by questions measurement method had a slightly higher validity than telephone/by alters,[20] but had the worst reliability of all three methods. The personal interview by alters measurement method had relatively good reliability, but the worst validity. Therefore, the telephone/by alters measurement method appears to be the optimal choice when measuring the characteristics of ties in egocentered networks.

The effects of network size were also consistent. It was found that respondents who had smaller social networks had, as expected, a higher validity of measurement. However, the effect of network size on the reliability of measurement was statistically non-significant, though there was a tendency for tie characteristics in smaller networks to be measured less reliably than in larger networks.

Some personal characteristics also had consistent effects on the quality of measurement. Older respondents had lower reliability (marginally significant) and lower validity (significant) of measurement values. Gender had a statistically significant effect only on the validity of measurement. Tie characteristics were, on average, more validly measured among males. The effects of education proved to be statistically non-significant.

Statistically significant effects were also produced by question type, extraversion and emotional stability, but only in terms of the validity of measurement. Behavioral questions, as compared to questions with emotional content, were measured with somewhat greater validity. A possible reason may be that the measurement scale may appear to be more exact, that is, when a behavioral type question is presented. Consistent with personality theory, those who were both more extraverted and emotionally stable had a higher validity of measurement.

The strength of the effect of each factor can also be summarized. With one exception, the strongest proved to be the effect of the measurement method on both reliability and validity. Network size had a weaker effect on the reliability of measurement than gender, age and question type. Age had a stronger effect on reliability than gender. On the other hand, it seemed that network size had a stronger effect on validity of measurement, since its effect was stronger than the effects of gender, age and question type. On the other hand, since some results regarding the effects of network size were contrary to our expectations (lower reliability in smaller networks), these effects should be studied further. The interaction effects of network size, age and gender should also be studied, an opportunity, which, in our case, was made impossible by the limited sample size.

[20] It may be because, in all cases, it is used in the second wave and research has shown (e.g. Scherpenzeel, 1995a; Ferligoj and Hlebec, 1998, 1999) that the method that is used in the second or later waves is always better, since respondents learn how to respond to a certain measurement instrument and therefore produce fewer errors.

Acknowledgements

The authors would like to thank Brendan Bunting and the anonymous reviewers for valuable comments on earlier drafts of the paper.

Appendix A. Extraversion and emotional stability scales

Response scale:

1. Very inaccurate
2. Moderately inaccurate
3. Neither inaccurate nor accurate
4. Moderately accurate
5. Very accurate

Extraversion

Am the life of the party.
Feel comfortable around people.
Start conversations.
Talk to a lot of different people at parties.
Don't mind being the center of attention.
Don't talk a lot.
Keep in the background.
Have little to say.
Don't like to draw attention to myself.
Am quiet around strangers.

Emotional stability

Am relaxed most of the time.
Seldom feel blue.
Get stressed out easily.
Worry about things.
Am easily disturbed.
Get upset easily.
Change my mood a lot.
Have frequent mood swings.
Get irritated easily.
Often feel blue.

Appendix B. An example of a data matrix for the meta analysis

Dependent variables		Predictors		
Reliability coefficient	Validity coefficient	Method	Gender	Age
.96	.94	1	1	1
.83	.99	2	1	1
.85	.99	3	1	1
.85	.90	1	1	1
.89	.99	2	1	1
.76	.99	3	1	1
.94	.91	1	1	1
.72	.99	2	1	1
.82	.99	3	1	1
.81	.98	1	1	2
.85	.99	2	1	2
.76	.99	3	1	2
.65	.92	1	1	2
.95	.99	2	1	2
.64	.99	3	1	2
.69	.92	1	1	2
.80	.99	2	1	2
.78	.99	3	1	2
.97	.99	1	2	1
.87	.98	2	2	1
.87	.97	3	2	1
.88	.94	1	2	1
.81	.97	2	2	1
.81	.95	3	2	1
.80	.93	1	2	1
.94	.98	2	2	1
.79	.96	3	2	1
.99	.96	1	2	2
.78	.92	2	2	2
.78	.96	3	2	2
.74	.87	1	2	2
.93	.92	2	2	2
.78	.94	3	2	2
.73	.89	1	2	2
.92	.93	2	2	2
.82	.96	3	2	2

Labels: method: (1) personal interview/by alters; (2) telephone interview/by alters; (3) telephone interview/by questions; gender: (1) male; (2) female; age: (1) up to 40 years; (2) 41 years or more.

References

Althauser, R.P., Heberlein, T.A., Scott, R.A., 1971. A causal assessment of validity: the augmented multitrait-multimethod matrix. In: Blalock Jr., H.M. (Ed.), Causal Models in the Social Sciences. Aldine, Chicago, pp. 151–169.

Alwin, D., 1974. An analytic comparison of four approaches to the interpretation of relationships in the multitrait-multimethod matrix. In: Costner, H.L. (Ed.), Sociological Methodology 1973–1974. Jossey-Bass, San Francisco, pp. 79–105.

Andrews, F.M., Morgan, J.N., Sonquist, J.A., Klem, L., 1973. Multiple classification analysis. In: A Report on a Computer Program for Multiple Regression Using Categorical Predictors. The Institute for Social Research, University of Michigan, Ann Arbor.

Antonucci, T.C., 1985. Social support: theoretical advances, recent findings and pressing issues. In: Sarason, I.G., Sarason, B.R. (Eds.), Social Support: Theory, Research and Applications. Martinus Nijhoff Publishers, Dordrecht, pp. 21–37.

Antonucci, T.C., Akiyama, H., Lansford, J.E., 1998. Negative effects of close social relations. Family Relations 47, 379–384.

Berkman, L.F., 1985. The relationship of social networks and social support to morbidity and mortality. In: Cohen, S., Syme, S.L. (Eds.), Social Support and Health. Academic Press, Orlando, pp. 241–262.

Bernard, H.R., Killworth, P.D., 1977. Informant accuracy in social network data II. Human Communications Research 4, 3–18.

Bernard, H.R., Killworth, P.D., Sailer, L., 1979/1980. Informant accuracy in social network data iv: a comparison of clique-level structure in behavioral and cognitive network data. Social Networks 2, 191–218.

Bernard, H.R., Killworth, P.D., Sailer, L., 1982. Informant accuracy in social network data in: an experimental attempt to predict actual communication from recall data. Social Science Research 11, 30–66.

Bernard, H.R., Killworth, P.D., Kronenfeld, D., Sailer, L., 1985. On the validity of retrospective data. Annual Review of Anthropology 13, 495–517.

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Brewer, D.D., 2000. Forgetting in the recall-based elicitation of personal and social networks. Social Networks 22, 29–43.

Browne, M.W., 1984. The decomposition of multitrait-multimethod matrices. British Journal of Mathematical and Statistical Psychology 37, 1–21.

Browne, M.W., 1985. MUTMUM, decomposition of multitrait-multimethod matrices. University of South Africa, Department of Statistics, Pretoria, South Africa.

Cambell, D.T., Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 56, 81–105.

Campbell, K.E., Marsden, P.V., Hurlbert, J.S., 1986. Social resources and socioeconomic status. Social Networks 8, 97–117.

Cantril, H., 1944. Gauging Public Opinion. Princeton University Press, Princeton.

Casciaro, T., 1998. Seeing things clearly: social structure, personality and accuracy in social network perception. Social Networks 20, 331–351.

Converse, J.M., Presser, S., 1988. Survey Questions: Handcrafting the Standardized Questionnaire. Sage, Beverly Hills.

Corman, S.R., Bradford, L., 1993. Situational effects on the accuracy of self-reported organizational communication. Communication Research 20, 799–817.

Eysenck, H.J., 1947. Dimensions of Personality. Routledge and Kegan Paul, London.

Eysenck, H.J., Eysenck, S.B.G., 1968. Manual for the Eysenck Personality Inventory. Educational and Industrial Testing Services, San Diego.

Ferligoj, A., Hlebec, V., 1995. Reliability of network measurements. In: Ferligoj, A., Kramberger, A. (Eds.), Contributions to Methodology and Statistics. FDV, Ljubljana, pp. 219–232.

Ferligoj, A., Hlebec, V., 1998. Quality of scales measuring complete social networks. In: Ferligoj, A. (Ed.), Advances in Methodology, Data Analysis, and Statistics, Metodološki zvezki, 14. FDV, Ljubljana, pp. 173–186.

Ferligoj, A., Hlebec, V., 1999. Evaluation of social network measurement instruments. Social Networks 21, 111–130.

Ferligoj, A., Leskošek, K., Kogovšek, T., 1995. Zanesljivost in veljavnost merjenja. FDV, Ljubljana.

Fischer, C.S., 1982. To Dwell Among Friends. University of Chicago Press, Chicago.

Freeman, L.C., Romney, A.K., 1987. Words, deeds and social structure: a preliminary study of the reliability of informants. Human Organizations 46, 330–334.

Freeman, L.C., Romney, A.K., Freeman, S.C., 1987. Cognitive structure and informant accuracy. American Anthropologist 89, 311–325.
Gottlieb, B.H., 1983. Social support strategies: guidelines for mental health practice. Sage, Beverly Hills.
Hall, C.S., Lindzey, G., Campbell, J.B. (Eds.), 1998. Theories of Personality, fourth ed. Wiley, New York.
Hammer, M., 1984. Explorations into the meaning of social network interview data. Social Networks 6, 341–371.
Hippler, H.J., Schwarz, N., Sudman, S. (Eds.), 1987. Social Information Processing and Survey Methodology. Springer-Verlag, New York.
Hirsch, B.J., 1981. Social networks and the coping process: creating personal communities. In: Gottlieb, B.H. (Ed.), Social Networks and Social Support. Sage, Beverly Hills, pp. 149–170.
Hlebec, V., 1993. Recall versus recognition: comparison of the two alternative procedures for collecting social network data. In: Ferligoj, A., Kramberger, A. (Eds.), Developments in Statistics and Methodology. FDV, Ljubljana, pp. 121–128.
Hlebec, V., 1999. Evaluation of Survey Measurement Instruments for Measuring Social Networks (doctoral dissertation). FDV, Ljubljana.
Hlebec, V., Ferligoj, A., 1996. Kvaliteta merjenja družbenih omrežij. In: Kramberger, A. (Ed.), Slovenska država, družba in javnost. FDV, Ljubljana, pp. 151–162.
Hlebec, V., Ferligoj, A., 2001. Respondent mood and the instability of survey network measurements. Social Networks 23, 125–140.
Hlebec, V., Ferligoj, A., 2002. Reliability of social network measurement instruments. Field Methods 14, 288–306.
Hooker, K., Monahan, D.J., Bowman, S.R., Frazier, L.D., Shifren, K., 1998. Personality counts for a lot: predictors of mental and physical health of spouse caregivers in two disease groups. The Journals of Gerontology 53B, P73–P85.
House, J.S., 1981. Work Stress and Social Support. Addison-Wesley, Reading.
Killworth, P.D., Bernard, H.R., 1976. Informant accuracy in social network data. Human Organization 3, 269–286.
Killworth, P.D., Bernard, H.R., 1979/1980. Informant accuracy in social network data III: a comparison of triadic structure in behavioral and cognitive data. Social Networks 2, 10–46.
Kogovšek, T., Ferligoj, A., Coenders, G., 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.
Költringer, R., 1995. Measurement quality in austrian personal interview surveys. In: Saris, W.E., Münnich, A. (Eds.), The Multitrait-Multimethod Approach to Evaluate Measurement Instruments. Eötvös University Press, Budapest, pp. 207–224.
Latkin, C., Mandell, W., Vlahov, D., Knowlton, A., Oziemkowska, M., Celentano, D., 1995. Personal network characteristics as antecedents to needle-sharing and shooting gallery attendance. Social Networks 17, 219–228.
Lord, F.M., Novick, M.R., 1968. Statistical Theories of Mental Test Scores. Addison-Wesley, Reading.
Marsden, P.V., 1987. Core discussion networks of americans. American Sociological Review 52, 122–131.
Marsden, P.V., Campbell, K.E., 1984. Measuring tie strength. Social Forces 63, 482–501.
Marsh, H.W., 1989. Confirmatory Factor Analysis of multitrait-multimethod data: many problems and few solutions. Applied Psychological Measurement 13, 335–361.
Matthews, G., Wells, A., 1999. The cognitive science of attention and emotion. In: Dalgleish, T., Power, M. (Eds.), Handbook of Cognitions and Emotion. Wiley, Chichester, pp. 171–192.
Mongrain, M., 1998. Parental representations and support-seeking behaviors related to dependency and self-criticism. Journal of Personality 66, 151–173.
Newcomb, M.D., Keefe, K., 1997. Social support, self-esteem, social conformity, and gregariousness: developmental patterns across 12 years. In: Pierce, G.R., Lakey, B., Sarason, I.G., Sarason, B.R. (Eds.), Sourcebook of Social Support and Personality. Plenum Press, New York, pp. 303–333.
Nolen-Hoeksema, S., Christopher, C.G., 1999. "Thanks for Sharing That": ruminators and their social support networks. Journal of Personality and Social Psychology 77, 801–814.
Payne, S.L., 1951. The Art of Asking Questions. Princeton University Press, Princeton.
van der Poel, M.G.M., 1993. Personal Networks. Swets and Zeitlinger, Lisse.
Procidano, M.E., Heller, K. (Eds.), 1983. Measures of perceived social support from friends and from family: three validation studies. American Journal of Community Psychology 11, 1–24.
Romney, A.K., Faust, K., 1982. Predicting the structure of a communications network from recalled data. Social Networks 4, 285–306.

Romney, A.K., Weller, S.C., 1984. Predicting informant accuracy from patterns of recall among individuals. Social Networks 4, 59–77.

Samuelsson, M.A.K., 1997. Social networks of children in single-parent families: differences according to sex, age, socioeconomic status and housing-type and their associations with behavioral disturbances. Social Networks 19, 113–127.

Sarason, I.G., Levine, H.M., Basham, R.B., Sarason, B.R., 1983. Assessing social support: the social support questionnaire. Journal of Personality and Social Psychology 44, 127–139.

Sarason, I.G., Sarason, B.R., 1985. Social support—insights from assessment and experimentation. In: Sarason, I.G., Sarason, B.R. (Eds.), Social Support: Theory, Research and Applications. Martinus Nijhoff Publishers, Dordrecht, pp. 39–50.

Saris, W.E., October 1999. Forced Choice or Agree/disagree questions? an evaluation by the split ballot MTMM experiment. In: Proceedings of the meeting of the IRMCS, pp. 122–146.

Saris, W.E., Andrews, F.M., 1991. Evaluation of measurement instruments using a structural modeling approach. In: Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (Eds.), Measurement Errors in Surveys. Wiley, New York, pp. 575–597.

Saris, W.E., Coenders, G., Satorra, A., 2001. A new approach for evaluating quality of measurement instruments: split ballot MTMM design. Working paper.

Saris, W.E., Meurs, A. Van (Eds.), 1990. Evaluation of measurement instruments by meta analysis of multitrait-multimethod studies. North-Holland, Amsterdam, pp. 134–147.

Saris, W.E., Münnich, A. (Eds.), 1995. The Multitrait-Multimethod Approach to Evaluate Measurement Instruments. Eötvös University Press, Budapest.

Scherpenzeel, A., 1995a. Meta Analysis of a European Comparative Study. In: Saris, W.E., Münnich, A. (Eds.), The Multitrait-Multimethod Approach to Evaluate Measurement Instruments. Eötvös University Press, Budapest, pp. 225–242.

Scherpenzeel, A., 1995b. Misspecification effects. In: Saris, W.E., Münnich, A. (Eds.), The Multitrait-Multimethod Approach to Evaluate Measurement Instruments. Eötvös University Press, Budapest, pp. 61–70.

Scherpenzeel, A., 1995c. A Question of Quality: Evaluating Survey Questions by Multitrait-Multimethod Studies. Royal PTT Netherlands, Amsterdam.

Schwarz, N., Sudman, S. (Eds.), 1992. Context Effects in Social and Psychological Research. Springer-Verlag, New York.

Schwarz, N., Sudman, S. (Eds.), 1994. Autobiographical Memory and the Validity of Retrospective Reports. Springer-Verlag, New York.

Schwarz, N., Sudman, S. (Eds.), 1996. Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. Jossey Bass, San Francisco.

Sudman, S., Bradburn, N.M., 1982. Asking Questions: A Practical Guide to Questionnaire Design. Jossey-Bass, San Francisco.

Sudman, S., Bradburn, N.M., Schwartz, N., 1996. Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. Jossey-Bass, San Francisco.

Turner, H.A., 1994. Gender and Social Support—Taking the Bad with the Good Sex Roles 30, 521–541.

Vinokur, A., Schul, Y., Caplan, R., 1987. Determinants of perceived social support: interpersonal transactions, personal outlook, and transient affective states. Journal of Personality and Social Psychology 53, 1137–1145.

Wellman, B., Wortley, S., 1990. Different strokes from different folks. American Journal of Sociology 96, 558–588.

Wenger, C.G., 1994. Support Networks of Older People: A Guide for Practitioners. Centre for Social Policy Research and Development, University College of Wales, Bangor.

Werts, C.E., Linn, R.L., 1970. Path analysis: psychological examples. Psychological Bulletin 74, 193–212.