Reliability of Social Network Measurement Instruments

Valentina Hlebec and Anuska Ferligoj

Field Methods 2002; 14; 288
DOI: 10.1177/15222X014003003

Reliability of Social Network Measurement Instruments

VALENTINA HLEBEC
ANUŠKA FERLIGOJ
University of Ljubljana

This article evaluates the quality of instruments for measuring support in social networks. The authors discuss the results of ten experiments designed to analyze the reliability of five measurement scales as well as two measurement methods for listing alters (free recall and recognition), type of network question (original, reciprocated), and characteristics of study design (time between instrument presentations). Analysis shows that the binary scale and the first presentation of measurement instruments are the least reliable. The most reliable were ordinal scales, among which the five-category ordinal scale with labels was the most reliable. The two data collection methods (free recall and recognition) and the two types of network questions (original, reciprocated) yield equally reliable data.

Survey data collection is the method used most often in social science research. Data collected by surveys have to be valid and reliable to allow statistical analysis, interpretation, and evaluation of results. Since the middle of the twentieth century when the first systematic analyses of questionnaire design appeared (Rugg and Cantrill 1944; Gallup 1947; Payne 1951), survey researchers have been concerned with questionnaire construction and its effects on data quality. Moreno (1953), the founding father of sociometry and later of social network analysis, was concerned with the design of appropriate measurement instruments; he particularly stressed the importance of the adjustment of sociometric measurement instruments to characteristics of the group being studied (Moreno 1953:44–45).

Much work concerning the quality of network data has been done on topics such as respondent accuracy (e.g., Killworth and Bernard 1976; Bernard, Killworth, and Sailer 1980; Bernard, Killworth, and Sailer 1982; Romney and Weller 1984; Romney, Weller, and Batchelder 1986; Brewer, Romney, and Batchelder 1991; Bondonio 1998; Casciaro 1998), characteristics of the measured networks (e.g., Burt 1984; Marsden 1987; Wellman and Wortley 1990), comparison of the measured networks using different network generators (e.g., Bernard, Shelley, and Killworth 1987; Killworth et al. 1990;

Field Methods, Vol. 14, No. 3, August 2002 288–306

Campbell and Lee 1991), and characteristics of the measured ties (e.g., Granovetter 1982; Marsden and Campbell 1984; Burt 1986). Owing to the complexity of the data, there is still need for an extensive and systematic evaluation of survey instruments in terms of quality of measurement in the field of social network analysis.

Some studies have focused mainly on the question of network data quality (Hammer 1984; Sudman 1985, 1988; Hlebec 1993; Brewer and Webster 1997; Ferligoj and Hlebec 1999). These articles are especially important for the design used in this study, which looks at the reliability of complete networks in which data are collected using respondents' reports. These studies indicated little difference (when the networks were relatively small) between the free-recall and recognition techniques in the assessment of (1) important ties, (2) most recent contacts, and (3) most frequent contacts. The recognition technique gave much better results when networks were larger and ties were weaker.

When evaluating measurement quality of complete networks, the percentage of mutually confirmed choices from two data sources (or data from the same data source taken twice, or the correlation of the complete network data matrices measured twice) is used as a measure of reliability. Other aspects of complete network data quality are also examined, such as validation of social positions (e.g., informal leadership [Macur and Hlebec 1996], centrality and measurement error [Ferligoj and Hlebec 1993], and cognitive processes in network measurements [Pattison 1994], among others).

Ferligoj and Hlebec (1995a, 1995b) analyzed the reliability of several measures of social support in a complete network. The true score multitrait-multimethod (MTMM) approach (Saris and Andrews 1991) on vectorized relational matrices was also used to evaluate the validity and reliability of four social support variables measured with three different scales (i.e., binary scale, eleven-category ordinal scale, and line-production scale). Social support measurements using binary scales were the least reliable. In a later study (Ferligoj and Hlebec 1999), the binary scale compared to five-category ordinal scales and the line-production method still gave the worst rating for reliability. The same is true for the first presentation of the measurement instrument.

In this article, we focus on the stability of the results of meta-analyses. Data from all previous studies (Ferligoj and Hlebec 1995a, 1995b, 1999) were combined to create a larger database suitable for meta-analysis. Apart from eight similar social networks (third-grade pupils from a grammar school—same age, similar size), there are two additional networks (a group of second-grade pupils from a vocational school and a group of second-year students from the University of Ljubljana). An additional scale—an

eleven-category ordinal scale—is used to evaluate the strength of social support flow. Our main goals were to find out (1) whether the results of meta-analyses are stable, regardless of variability in the size of the social networks included or the age of respondents, and (2) how reliable the eleven-category ordinal scale is.

HYPOTHESES

Hypothesis 1: The stability of social support provisions should remain similar to that found in previous meta-analysis (Ferligoj and Hlebec 1995a, 1995b, 1998, 1999; Hlebec 1999), regardless of additional social networks included (i.e., the size and characteristics of respondents should not affect the reliability of social support dimensions). Therefore, emotional support should be the most stable, as it is provided by close and important ties. Material support should be the least reliable, as it is provided by specialized sources. Closeness is not required for providers of material support. The same is true for providers of informational support. However, the providers of informational support in our study are called on in cases of longer illness. Therefore, this resource is provided by more important ties.

Hypothesis 2: As shown in previous studies (Ferligoj and Hlebec 1995a, 1995b, 1998, 1999; Hlebec 1999), the ordinal scales are favored among Slovenian youth. Therefore, the five-category ordinal scale should be one of the most stable scales. The binary scale should remain the least stable.

Hypothesis 3: The size of the social network and age of the respondent should not affect the similarity of free-recall and recognition methods. Therefore, results should be similar to those in previous meta-analyses.

Hypothesis 4: The measurement instrument first presented to respondents should have the lowest reliability, as in previous studies. Network size and age of respondents should not interact with presentation ordering with regard to reliability.

MEASURES

In empirical research, social support is sometimes used as a dependent variable and sometimes as a predictor variable that can influence outcomes such as social isolation, depression, stress, and health problems, among others (for a review, see Vaux 1988; Sarason, Sarason, and Gurung 1997). Results obtained from research are sometimes used to design social interventions (Rook and Dooley 1985; Vaux 1998; Sarason, Sarason, and Pierce 1990; Duck 1997), so different definitions of social support can have real consequences.

Different authors have formulated various classifications of support activities (for an overview, see Vaux 1988). Cohen and Wills (1985) have proposed a useful typology of support provisions. They compared more than fifty empirical studies of social support and proposed the following four types of social support. (1) Esteem support (emotional or close support) comprises information that a person is accepted and esteemed. (2) Informational support (advice support, appraisal support, or cognitive guidance) involves help in defining, understanding, and coping with problematic events. (3) Social companionship (diffuse support and belongingness) relates to time spent with others in leisure activities. (4) Instrumental support (aid, material support, or tangible support) is the provision of financial aid, material support, and necessary services. The typology Cohen and Wills presented has been used in many empirical studies, and the components of the typology have also been identified in children (Cauce et al. 1990).

Cohen and Wills's (1985) typology was the basis for determining the number and types of social support provisions we measured in our study. We used several criteria in designing the four network generators that measured emotional support, instrumental support, social companionship, and informational support. We sought appropriate network generators from a pool of previously known and tested network generators to ensure the comparability and usefulness of our results. The decision to use only one network generator to measure a particular dimension of social support was based on time limitations. School authorities allowed only two interviews for the data collection in the grammar school. Therefore, the measurement instrument had to be short enough to ensure the three measurements necessary for the MTMM design.

On the other hand, name generators had to be adjusted to the characteristics of the observed groups (i.e., classrooms, as strongly recommended by Moreno 1953:44–45). Eight groups were selected from third-year classrooms in a grammar school in Ljubljana, the capital of Slovenia. On average, there were thirty-one students, aged seventeen, in each of eight classes. One group was selected from a vocational school in Ljubljana; the average age of these students was fifteen. There were thirty-four students in this group. One group was selected from the Faculty of Social Sciences in Ljubljana. This group was smaller (thirteen members), and the average age was twenty-one.

Using the criteria described above, the following network generators were used. (1) Material support was measured by exchange of study materials (the exact wording is presented in the appendix), in case of absences. This network generator measures the exchange of small services or material support from specialized providers who are not necessarily personally important to

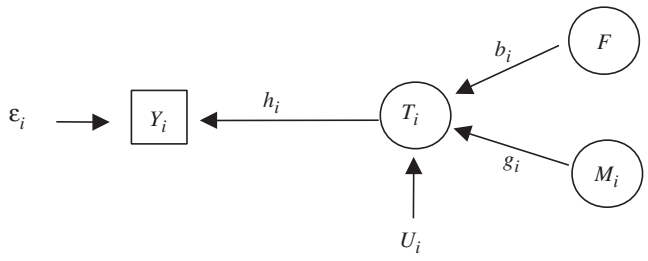
the respondent but who have the best notes. (2) Informational support was assessed from information about important study assignments in the case of a hypothetical severe illness that would require hospitalization in May, the most important grading month. This service should be provided by close friends or trustworthy others who are geographically close. (3) Social companionship was measured by invitation to a hypothetical birthday party proposed as taking place a week after interviewing. The selection of this stimulus was quite difficult, as an activity had to be selected in which every student could be involved. Thus, extracurricular activities would not have been appropriate. (4) Burt's (1984) network generator was used to measure emotional support exchange. The discussion of important personal matters was not limited to a period of the past six months, as in Burt's original measure. Instead of evaluation based on a recent exchange, evaluation of an underlying pattern was required, based on the preference of respondents themselves. Emotional support should be provided by very close and important others.

The reciprocity of social support and the balance of social support exchange are important sources of well-being (Antonucci and Jackson 1990; Goodenow, Reisine, and Grady 1990; Jung 1997). Therefore, both directions of social support exchange were assessed. All the name generators were applied twice. First, respondents were asked to describe whom they would ask for a particular exchange (original question); then, respondents were asked who would ask them for a particular exchange (reversed question). Social support was thus measured in giving and receiving.

As shown by Marsden and Campbell (1984), the best measure of relationship strength is individual evaluation of a relationship's closeness and importance. To measure the intensity of social support provision, we designed several measurement scales to evaluate the strength of social support provision as perceived by the respondent. The scales assessed the likelihood that social support provision would take place. We selected a binary scale that indicates only the presence or absence of a relationship since this scale is used most often in social network analysis. It is simple to apply, but the measurement is quite crude. We also used four other scales.

The selection of scales was based on the findings of previous meta-analyses of survey data quality (for a review, see Hlebec 1999:59–69) and on analyses of complete network data quality (Ferligoj and Hlebec 1995a, 1995b, 1998, 1999). The most consistent finding of these studies was the strong impact of measurement scales on data quality. To measure the strength of relationships, we used five measurement scales (see the description of the scales in the appendix): (1) a binary scale, (2) a five-category ordinal scale with labeled extreme values, (3) a five-category ordinal scale with

FIGURE 1
True Score Measurement Model



all values labeled, (4) a line-drawing scale, and (5) an eleven-category ordinal scale.

In most, if not all data collection modes for collecting network data (in which respondents are the source of information), two general approaches can be distinguished. In the first approach, the recognition method, respondents are presented with a list of all members of the group and are asked to estimate the strength of their relationships with each listed person. In the second approach, the free-recall format, respondents are not offered any help in selecting the names of significant others. We used both approaches to see whether the quality of data is significantly better when respondents are given a roster, a technique often assumed to facilitate responses. Free recall and recognition are not directly compared within groups (i.e., the data collection technique was not varied within class). These two data collection techniques are compared in the meta-analysis.

ESTIMATING RELIABILITY

To estimate reliability, we applied the MTMM approach proposed by Campbell and Fiske (1959). There are several ways to assess the MTMM matrix. The true score measurement model, as proposed by Saris and Andrews (1991:576–83), in which reliability and validity (as well as the effect of the measurement method on the variance of true score) can be estimated, was selected. The true score measurement model is presented in Figure 1.

In this measurement model, Y_i is the response or observed variable (network generator) corresponding to the question measured by the method i; T_i

is the stable component when the same question is repeated under exactly the same conditions; ε_i is the random error in the observed variable Y_i; F is the unobserved variable of interest, assumed to be independent of the measurement procedure used (social support provision); M_i is a method-specific component; and U_i is the unique component of the true score, T_i. In this model, it is assumed that:

$$E(\varepsilon_i) = 0;\ E(U_i) = 0;\ cov(F, U_i) = 0;\ cov(M_i, U_i) = 0;\ cov(M_i, \varepsilon_i) = 0;$$
$$cov(F, \varepsilon_i) = 0;\ cov(U_i, \varepsilon_i) = 0;\ cov(F_i, M_i) = 0.$$

In this model, one can estimate reliability (h_i), the true score validity (b_i), and method effects (g_i) for a single variable, which is a unique feature of this model. Reliability is defined as the proportion of the variance in Y_i that remains stable across repetitions of the same measure, or in this measurement model:

$$reliability = \frac{var(T_i)}{var(Y_i)} = h_i^2.$$

STUDY DESIGN

We applied only three scales in each class, in keeping with a traditional MTMM design. Within each class, we varied the ordering of three selected scales and the time intervals between three repetitions. Across classes, we varied data collection techniques and compositions of selected scales. The outline of the study design is presented in Table 1.

As Table 1 shows, there were ten classes. In the first six classes, we used recognition data collection; in the last four classes, we used the free-recall data collection technique. In the first class, we used the binary scale, the line-production scale, and an eleven-category ordinal measurement scale in the described ordering. We included all presentations of the network measurement instrument within one interview at intervals of approximately twenty minutes. We carried out the data collection in May 1993. The same information can be obtained for other classes from Table 1. We used the paper-and-pencil data collection mode in all groups. The first group is made up of university students, the second group is made up of pupils from a vocational school, and the remaining eight groups are made up of pupils from a grammar school.

TABLE 1
Study Design

Group	*Scale*	*Ordering*	*Interview*	*Data Collection Method*	*Date of Survey*
1	1	1	1	1	May 1993
	4	2	1	1	May 1993
	5	3	1	1	May 1993
2	1	3	3	1	May 1995
	4	2	2	1	May 1995
	5	1	1	1	May 1995
3.1	1	1	2	1	January 1998
	2	2	2	1	January 1998
	3	3	1	1	January 1998
3.2	1	2	2	1	January 1998
	2	3	2	1	January 1998
	4	1	1	1	January 1998
3.3	1	3	1	1	January 1998
	3	1	2	1	January 1998
	4	2	2	1	January 1998
3.4	2	1	1	1	January 1998
	3	2	2	1	January 1998
	4	3	2	1	January 1998
3.5	1	1	1	2	January 1998
	2	2	2	2	January 1998
	3	3	2	2	January 1998
3.6	1	2	2	2	January 1998
	2	3	1	2	January 1998
	4	1	2	2	January 1998
3.7	1	1	2	2	January 1998
	3	2	1	2	January 1998
	4	3	2	2	January 1998
3.8	2	1	2	2	January 1998
	3	2	2	2	January 1998
	4	3	1	2	January 1998

NOTE: Scale: 1 = binary scale; 2 = ordinal scale (five category, extreme values labeled); 3 = ordinal scale (five category, all values labeled); 4 = line-production scale; 5 = ordinal scale (eleven category, extreme values labeled). Ordering: 1 = first; 2 = second; 3 = third. Interview: 1 = first interview; 2 = second interview; 3 = third interview. Data collection method: 1 = recognition; 2 = free recall.

META-ANALYSIS

Once data sets across all classes were collected, we vectorized each of the twelve relation matrices (four dimensions of social support × three measurement scales) for each class. Then we estimated the reliability coefficients for

each of the vectorized relational matrices within each of the ten classes. In the last phase, we performed a meta-analysis on reliability estimates in which the variability of reliability coefficients was explained by characteristics of the measurement instruments.

The analysis of the MTMM model was based on a matrix of Pearson's correlation coefficients. The reliability coefficients were obtained by the maximum likelihood procedure in the LISREL VI program (Jöreskog and Sörbom 1986), using the true score MTMM model. For each group, we constructed two MTMM matrices, one for the original four questions and the other for the reversed four questions measuring social support (i.e., giving and receiving).

Following the example given in other meta-analyses for explaining the effects on the data-quality estimates of different characteristics in the measurement instruments (Scherpenzeel 1995; Ferligoj and Hlebec 1999), we chose multiple classification analysis (MCA) as the meta-analysis technique (Andrews et al. 1973). The multivariate MCA coefficients indicate how much the reliability estimates deviate from the mean as a result of a given characteristic of the measurement instrument, while controlling for the effects of all other characteristics of the measurement instrument. Two measures of the overall effect of each predictor are obtained; the MCA η and MCA β are also obtained.

The MCA η coefficient measures the strength of the bivariate relationship between a quality estimate and a predictor. MCA β coefficients, on the other hand, measure the strength of the relationship, controlled for the other predictor variables in the model. The rank order of the βs indicates the relative importance of the predictor variables in their explanation of the dependent variable. Finally, the multiple R^2, indicating the total proportion of variance explained by all predictors together, is estimated.

The reliability coefficients from the MTMM models were used as the dependent variables in a meta-analysis. The predictor variables in meta-analysis are the characteristics of the measurement instruments and the characteristics of the true score MTMM approach to estimating measurement quality. Each measurement instrument is characterized by the type of social support measured, its response scale, type of data collection method, and the characteristics of the study in which it was included (i.e., time between presentations).

RESULTS

Results of two separate meta-analyses are presented in this section. Two separate analyses were needed, as there are several predictive variables and a

TABLE 2
Predictive Power and Effects for the Social Support Domain, Multitrait-Multimethod Design, Data Collection Method, and Type of Question on Reliability Estimates

		Reliability Coefficient (M = .879) Multivariate Measures			
	n	η	β	*Deviation*	*Reliability*
Social support					
Material	60			–.016	.863
Informational	60			.010	.889
Social companionship	48			–.001	.878
Emotional	48	.180	.178	.009	.888
Multitrait-multimethod design					
First presentation	72			–.032	.847
Repeated after twenty minutes	72			.042	.921
Repeated after one week	72	.522	.522	–.010	.869
Data collection method					
Recognition	120			–.003	.876
Free recall	96	.053	.047	.003	.882
Type of question					
Original	108			.002	.881
Reciprocated	108	.032	.032	–.002	.877
Multiple R^2			.308		

smaller number of cases, especially for the eleven-category ordinal scale. Results in Table 2 show the effects for the type of social support, characteristic of MTMM design, data collection method, and type of question on the reliability estimates.

The mean reliability is .879 (minimum .71, maximum .99), which is reasonably high. The standard deviation is .06, which is small but expected and comparable to similar studies done on attitudinal variables in several European countries (see, e.g., Scherpenzeel 1995). The multiple R^2 in Table 2 shows that these four predictors explain 31% of the variance of the reliability estimates. The largest multivariate effects for the reliability estimates occur in characteristic of MTMM design and social support domain (βs are .552 and .180, respectively). In contrast, data collection method and type of question do not seem to have much effect on the reliability estimates. Deviations from the mean reliability show to what extent reliability deviates from the grand mean as the result of a particular predictor variable. In other words, if

TABLE 3
Predictive Power and Effects for the Measurement Scales, Multitrait-Multimethod Design, and Type of Question on Reliability Estimates

		Reliability Coefficients (M = .879) Multivariate Measures			
	n	η	β	*Deviation*	*Reliability*
Measurement scale					
Binary scale	56			–.044	.835
Five-category ordinal scale	48			.015	.894
Line-production scale	56			.006	.885
Five-category ordinal scale with labeled answers	48			.025	.904
Eleven-category ordinal scale	8	.453	.446	.022	.901
Type of question					
Original	108			.002	.881
Reciprocated	108	.032	.032	–.002	.877
Multitrait-multimethod design					
First presentation	72			–.032	.847
Repeated presentations	144	.380	.372	.016	.895
Multiple R^2			.344		

the question about social support is presented first, its reliability decreases by .032 (.879 – .032 = .847).

As hypothesized, the first measurement is the least stable. As expected, emotional support and informational support are measured more reliably than material support or social companionship. The results obtained show that results of previous studies can be generalized to social networks of various sizes and respondents of various ages. Data collection method and type of question do not affect reliability. In networks of thirteen to thirty-four members, it is irrelevant whether social support is evaluated using a roster, at least on the level of the network as whole. Both given support and received support are evaluated equally reliably.

Results of the second meta-analysis are shown in Table 3. Here, the effects for the measurement scales, MTMM design, and type of question on reliability estimates are presented.

To prevent empty cells in the classification table, variable MTMM design had to be recoded into two categories. There were only two cases in which the eleven-category ordinal scale was used. Therefore, only the first measure-

ment of social support provision and the repeated measurement are distinguished. The multiple R^2 shows that the three predictors above explain 34% of the variance of the reliability estimates. The largest multivariate effects for the reliability estimates occur in type of measurement scale and characteristic of MTMM design (βs are .446 and .372, respectively). The binary scale is still the least reliable. The five-category ordinal scale with all categories labeled is still the most reliable. The eleven-category ordinal scale with labeled extremes is the second most reliable (even more reliable than the five-category ordinal scale with labeled extremes). It seems that when the number of categories is small, Slovenian respondents prefer labeling all categories. However, when only extremes are labeled, a large number of categories seems more reliable.

The first measurement of social support is still the least reliable when compared to repeated measurements. For inexperienced respondents, reports about their relations and social support provision seem rather a difficult task. As there is usually only one measurement available (time and cost constraints), a rehearsal of response task is strongly recommended.

CONCLUSIONS

Formal aspects of the formulation of name generators are as important as their contents. Selection of proper data collection methods, measurement scales, question ordering, or graphic design can substantially increase measurement quality of measured social networks. As in meta-analyses, the reliability of survey-measured network data can be substantially improved when using an appropriate measurement instrument.

If one is interested in measuring social support provision in a smaller social network, then material support is measured with a mean reliability of .863 (.879 – .016). When material support provision is measured first without an example (–.032), using the recognition data collection technique (–.003), original question (+.002), and binary scale (–.044), its reliability is substantially lower (.786). When emotional support is measured in such a network, its mean reliability is .888 (.879 + .009). When emotional support is evaluated shortly after an example (+.042), using the free-recall technique (+.003), original question (+.002), and five-category ordinal scale with labeled categories (+.025), its reliability should be, and is, considerably higher (.960).

Among the measures of social support, the exchange of study materials is the least reliable measure. Apparently, the instrumental dimension of social support is less reliable because of its importance to the respondents and the characteristics of ties that provide material support. The social companion-

ship domain differs little from average reliability, whereas informational support and emotional support are measured with the highest reliability. Social support provided by strong ties is measured more reliably than is social support provided by weak ties.

The response scale is the second most important predictor of reliability estimates across both meta-analyses. It appears that the binary scale is the least reliable. This agrees with previous results (Ferligoj and Hlebec 1995a, 1995b, 1998, 1999). The five-category ordinal scale with labels is the most reliable, and the eleven-category ordinal scale is the second most reliable. The five-category ordinal scale without labels is still better than the line-drawing scale.

In previous research (Ferligoj and Hlebec 1995a, 1995b, 1998), the binary scale was compared to both the eleven-category ordinal scale and the line-drawing scale. In this combination, both the ordinal scale and the line-drawing scale were equally good and much better than the binary scale. In previous meta-analyses (Ferligoj and Hlebec 1999), the binary scale was compared to the line-drawing scale and two five-category ordinal scales. In this combination, both ordinal scales were most reliable. All five measurement scales are put together here for the first time. It appears that in Slovenia, respondents accept ordinal scales better than other types of scales.

When all the scales are put together, the line-production scale is estimated to be less reliable when compared to five-category and eleven-category ordinal scales. This result differs from previous analyses, in which the line-production scale was compared only to the binary scale and the eleven-category ordinal scale. Results of meta-analyses done in this study cannot be treated as conclusive, as only a small number of cases involving the eleven-category ordinal scale were included. Further analyses should reveal whether, in social networks of high school students, the line-production scale is less appropriate for measuring social support than ordinal scales in general.

The data collection method does not have much effect on the quality estimates. It appears that for the type of relationship in which respondents know each other very well, free recall functions just as well as the method in which the full list of members is presented, when stability of measurement is in question. This is true for social support and for some other types of relationships (see Hammer 1984; Sudman 1985, 1988). Nevertheless, when one uses the recognition data collection technique, more ties and weaker ties are also reported, in contrast to reports from the free-recall technique.

Results of this meta-analysis show that a discrepancy between original and reversed questions is not that important in smaller networks, as has been

already suggested by previous meta-analyses (Ferligoj and Hlebec 1999). One can conclude only that the perception of social support expected from respondents is as stable as the social support needed by respondents.

The time between repetitions is the most important predictor variable in the first meta-analysis. When a measure is presented first, it is the least reliable. When a measure is repeated after twenty minutes, its reliability estimate significantly increases. When a measure is repeated after one week, its reliability is lower than the mean reliability. It is clear that the first presented measure has the lowest reliability estimates and that a short time between repetitions increases reliability. The same is true when the first presentation is distinguished from repeated presentations, as was shown in the second meta-analysis.

Compared to other meta-analyses (Ferligoj and Hlebec 1999) in which the observed groups were very similar (the same age of respondents, similar network size, and groups of classes within the same high school), results are rather stable regardless of the inclusion of different groups (age, network size, and intellectual abilities).

A high quality for measured complete networks was obtained at the general level, that is, at the level of the whole network. Further explorations are needed at the level of individual respondents, and other predictor variables, such as other measurement scales, should be included in experimental designs.

APPENDIX

Altogether there were eight different forms of network generators with varying scales and data collection techniques. The questionnaire had four sections:

1. Network generator measuring instrumental support with a binary scale and with the recognition data collection technique (original question):

> You have known your classmates for some time now. It sometimes happens that you cannot take courses for various reasons. From which of your classmates would you borrow study materials? Indicate your answers on the list below in the following way:
>
> Mark 1 in the box next to a person's name if you would borrow study material from her/him. Mark 0 in the box next to a person's name if you would not borrow study materials from her/him.

Reversed question:

> Which of your classmates would ask you to lend your study materials?

Instructions for respondents were the same as for the original question.

2. Network generator measuring informational support with an ordinal scale without labels and with the recognition data collection technique (original question):

> Suppose you were ill at the beginning of May and you had to stay in the hospital for a month. Which of your classmates would you ask to obtain information about important study assignments? Indicate your answers on the list below in the following way:
>
> Select a number from 0 to 4 (10) to indicate how likely you would be to ask your classmates for help. Mark 4 (10) in the box next to a person's name if you would certainly ask for help from her/him. Mark 0 in the box next to a person's name if you would not ask for help from her/him. The more likely it is that you would ask for help from a person, the larger the number should be.

Reversed question:

> Which of your classmates would ask you to obtain study information in the case of a long absence?

Instructions for respondents were the same as for the original question.

3. Network generator measuring companionship with a line production scale and with the free-recall data collection technique (original question):

> Suppose your birthday falls next week, and you want to give a birthday party. Which of your classmates would you invite? Indicate your answers on the list below in the following way:
>
> List the names of any classmates that you would invite to your birthday party; for each listed person, indicate by the length of the line how likely you would be to invite her/him. The longer the line, the more likely you would be to invite that person.

Reversed question:

> Which of your classmates would invite you to her/his birthday party?

Instructions for respondents were the same as for the original question.

4. Network generator measuring emotional help with an ordinal scale with labels and with the free-recall data collection technique (original question):

> With which of your classmates would you discuss important things? Indicate your answers on the list below in the following way.

List the names of any classmates with whom you would discuss important matters; for each listed person, use a number from 0 to 4 to indicate how likely you would be to discuss your important personal matters with her/him. Mark 4 if it is certain that you would discuss personal matters with her/him, mark 3 if it is very likely that you would discuss personal matters with her/him, mark 2 if it is likely that you would discuss personal matters with her/him, mark 1 if it is not likely that you would discuss personal matters with her/him, mark 0 if it is certain that you would not discuss personal matters with her/him.

Reversed question:

Which of your classmates would discuss important personal matters with you?

Instructions for respondents were the same as for the original question.

REFERENCES

Andrews, F. M., J. N. Morgan, J. A. Sonquist, and L. Klem. 1973. *Multiple classifications analysis*. Ann Arbor, MI: Institute for Social Research.

Antonucci, T. C., and J. S. Jackson. 1990. The role of reciprocity in social support. In *Social support: An interactional view*, edited by B. R. Sarason, I. G. Sarason, and G. R. Pierce, 173–98. New York: John Wiley.

Bernard, H. R., P. D. Killworth, and L. Sailer. 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.

———. 1982. Informant accuracy in social network data V: An experimental attempt to predict actual communication from recall data. *Social Science Research* 11:30–66.

Bernard, H. R., G. A. Shelley, and P. D. Killworth. 1987. How much of a network does the GSS and RSW dredge up? *Social Networks* 9:49–61.

Bondonio, D. 1998. Predictors of accuracy in perceiving informal social network. *Social Networks* 20:301–30.

Brewer, D. D., A. K. Romney, and W. H. Batchelder. 1991. Consistency and consensus: A replication. *Journal of Qualitative Anthropology* 3:195–205.

Brewer, D. D., and C. M. Webster. 1997. Recall and recognition of friends by residents of a university dormitory. Paper presented at the Seventeenth International Sunbelt Social Network Conference, San Diego, California.

Burt, R. S. 1984. Network items and the general social survey. *Social Networks* 6:293–339.

———. 1986. A note on sociometric order in the General Social Survey network data. *Social Networks* 8:149–74.

Campbell, D. T., and D. W. Fiske. 1959. Convergent and discriminant validation by the multitrait multimethod matrices. *Psychological Bulletin* 56:81–105.

Campbell, K. E., and A. L. Lee. 1991. Name generators in surveys of personal networks. *Social Networks* 13:203–21.

Casciaro, T. 1998. Seeing things clearly: Social structure, personality, and accuracy in social network perception. *Social Networks* 20 (4): 331–54.

Cauce, A. M., M. Reid, S. Landesman, and N. Gonzales. 1990. Social support in young children: Measurement, structure, and behavioral impact. In *Social support: An interactional view*, edited by B. R. Sarason, I. G. Sarason, and G. R. Pierce, 64–94. New York: John Wiley.

Cohen, S., and T. A. Wills. 1985. Stress, social support, and the buffering hypothesis. *Psychological Bulletin* 98 (2): 310–57.

Duck, S. 1997. *Handbook of communication and emotion: Research, theory, applications and contexts*. 2d ed. Chichester, UK: John Wiley.

Ferligoj, A., and V. Hlebec. 1993. (In)sensitivity of centrality measurements in social networks. In *Developments in statistics and methodology, Metodološki zvezki*, edited by A. Ferligoj and T. Kramberger, Vol. 9, 139–45. Ljubljana, Slovenia: FDV.

———. 1995a. Evaluating network data quality: An application of MTMM approach. In *The multitrait-multimethod approach to evaluate measurement instruments*, edited by W. Saris and Á Münnich, 155–72. Budapest, Hungary: Eötvös University Press.

———. 1995b. Reliability of network measurements. In *Contributions to methodology and statistics, Metodološki zvezk*, edited by A. Ferligoj and T. Kramberger, Vol. 10, 219–32. Ljubljana, Slovenia: FDV.

———. 1998. Quality of scales measuring complete social networks. In *Advances in methodology, data analysis and statistics, Metodološki zvezk*, edited by A. Ferligoj, Vol. 14, 173–86. Ljubljana, Slovenia: FDV.

———. 1999. Evaluation of social network measurement instruments. *Social Networks* 21:111–30.

Gallup, G. 1947. The quintamensional plan of question design. *Public Opinion Quarterly* 11:385–93.

Goodenow, C., S. T. Reisine, and K. E. Grady. 1990. Quality of social support and associated social and psychological functioning in women with rheumatoid arthritis. *Health Psychology* 9 (3): 266–84.

Granovetter, M. 1982. The strength of weak ties. In *Social structure and network analysis*, edited by P. V. Marsden and N. Lin, 105–30. Beverly Hills, CA: Sage.

Hammer, M. 1984. Explorations into the meaning of social network interview data. *Social Networks* 6:341–71.

Hlebec, V. 1993. Recall versus recognition: Comparison of two alternative procedures for collecting social network data. In *Developments in statistics and methodology, Metodološki zvezki*, edited by A. Ferligoj and T. Kramberger, Vol. 9, 121–28. Ljubljana, Slovenia: FDV.

———. 1999. *Evaluation of survey measurement instruments for measuring social networks*. Lubljana, Slovenia: FDV.

Jöreskog, K. G., and D. Sörbom. 1986. *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Uppsala, Sweden: University of Uppsala.

Jung, J. 1997. Balance and source of social support in relation to well-being. *Journal of General Psychology* 124 (1): 77–90.

Killworth, P. D., and H. R. Bernard. 1976. Informant accuracy in social network data. *Human Organization* 35 (3): 269–86.

Killworth, P. D., H. R. Bernard, C. McCarty, G. A. Shelley, and S. Robinson 1990. Comparing four different methods for measuring personal social networks. *Social Networks* 12:179–215.

Macur, M., and V. Hlebec. 1996. A combined qualitative and quantitative approach for analyzing the informal leadership of a choral society. In *Developments in data analysis,*

Metodološki zvezki, edited by A. Ferligoj and T. Kramberger, Vol. 12, 185–94. Ljubljana, Slovenia: FDV.

Marsden, P. V. 1987. Core discussion networks of Americans. *American Sociological Review* 52:122–31.

Marsden, P. V., and K. E. Campbell. 1984. Measuring ties strength. *Social Forces* 63 (2): 482–501.

Moreno, J. L. 1953. *Who shall survive? Foundations of sociometry, group psychotherapy and sociodrama*. New York: Beacon House.

Pattison, P. 1994. Social cognition in context. In *Advances in social network analysis: Research in the social and behavioral sciences*, edited by S. Wasserman and J. Galaskiewicz, 79–109. Thousand Oaks, CA: Sage.

Payne, S. L. 1951. *The art of asking questions*. Princeton, NJ: Princeton University Press.

Romney, A. K., and S. C. Weller. 1984. Predicting informant accuracy from patterns of recall among individuals. *Social Networks* 6:59–77.

Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* 88:313–38.

Rook, K. S., and D. Dooley. 1985. Applying social support research: Theoretical problems and future directions. *Journal of Social Issues* 41 (1): 5–28.

Rugg, D., and H. Cantril. 1944. The wording of questions. In *Gauging public opinion*, edited by H. Cantril et al., 23–50. Princeton, NJ: Princeton University Press.

Sarason, B. R., I. G. Sarason, and R.A.R. Gurung. 1997. Close personal relationships and health outcomes: A key to the role of social support. In *Handbook of communication and emotion: Research, theory, applications and contexts*, edited by S. Duck, 2d ed., 547–74. Chichester, UK: John Wiley.

Sarason, B. R., I. G. Sarason, and G. R. Pierce. 1990. *Social support: An interactional view*. New York: John Wiley.

Saris, W. E., and F. M. Andrews. 1991. Evaluation of measurement instruments using a structural modeling approach. In *Measurement errors in surveys*, edited by P. P. Biemer et al., 575–99. New York: John Wiley.

Scherpenzeel, A. 1995. A question of quality: Evaluating survey questions by multitrait-multimethod studies. Amsterdam: Royal PTT.

Sudman, S. 1985. Experiments in the measurement of the size of social networks. *Social Networks* 7:127–51.

———. 1988. Experiments in measuring neighbor and relative social networks. *Social Networks* 10:93–108.

Vaux, A. 1988. *Social support: Theory, research, and intervention*. New York: Praeger.

Wellman, B., and S. Wortley. 1990. Different strokes from different folks: Community ties and social support. *American Journal of Sociology* 96 (3): 558–88.

VALENTINA HLEBEC is an assistant professor of statistics at the University of Ljubljana. She has approximately twenty publications in social network analysis, questionnaire design, the sociology of adolescents and elderly, social psychology, and education. Some recent publications are "Evaluation of Social Network Measurement Instruments" (with A. Ferligoj, Social Networks*, 1999), "Sexual Satisfaction among Slovenian Youth" (with I. Bernik,* Drustvena Istrazivanja*, 2000), and "Respondent Mood and the Instability of Survey Network Measurements" (with A. Ferligoj,* Social Networks*, 2001).*

ANUŠKA FERLIGOJ is a professor of statistics at the University of Ljubljana and head of the Center of Methodology and Informatics and the Institute of Social Sciences in Ljubljana. She has approximately one hundred publications, including books she has edited or coedited on cluster analysis, research methodology, and network analysis, with applications in social sciences, economy, epidemiology, and education. Recent publications include "Symmetric-Acyclic Decompositions of Networks" (with P. Doreian and V. Batagelj, Journal of Classification*, 2000), "Consensus Decision Trees: Using Consensus Hierarchical Clustering for Data Relabeling and Reduction" (with B. Kavšek, and N. Lavrac,* Machine Learning*, 2001), and "Estimating the Reliability and Validity of Personal Support Measures: Full Information ML Estimation with Planned Incomplete Data" (with T. Kogovšek and and W. S. Saris,* Social Networks*, 2002).*