Mode Effects in Structural Modeling; A LISREL Multi-Group Comparison of Mail, Telephone, and Face to Face Survey Data

Edith D. de Leeuw und Joop J. Hox

1 Introduction

Until the 1970s, the face to face interview was the accepted method for conducting social science surveys. Since then there has been dramatic progress in other survey modes, primarily telephone and mail procedures (Dillman, 1978; Groves, 1989). This can be attributed to two important factors. First, the fast growing costs of the face to face interview and the growing non-response rates for this method have led survey researchers to consider alternative data collection procedures (Dillman, 1978; Dillman/Tarnai, 1988; Groves/Kahn, 1979; Goyder, 1985). Second, research in the last two decades shows that mail and telephone surveys have far greater potential than had previously been thought and that high response rates can be attained with acceptable survey costs and time constraints (Baumgartner/Heberlein, 1984; Dillman, 1991; Lyberg/Kasprzyk, 1991).

Nevertheless, acceptance of alternatives for the face to face interview has been limited, pending further demonstrations that the data quality would not suffer. As a result, the influence of data collection method on data quality has received considerable attention in survey research. A recent meta analysis of the research literature found that face to face interviews resulted in somewhat better overall data quality (e.g., lesser item nonresponse, more statements to open questions) when compared with telephone interviews (De Leeuw/Van der Zouwen, 1988). Self-administered questionnaires in comparison to both face to face and telephone interviews resulted in somewhat better data (i.e., less social desirability, more self disclosure), especially when sensitive questions are

In: Jost Reinecke & Gaby Krekeler (Hrsg.). (1993).*Methodische Grundlagen und Anwendungen von Strukturgleichungsmodellen. Band 1.* Mannheim: FRG e.V.

asked; but compared to both interview methods self-administered questionnaires also resulted in more item nonresponse (De Leeuw/Hox/Van der Zouwen, 1989).

In the studies reviewed, comparisons were mainly restricted to the analysis of univariate distributions, and little attention has been given to the possible effect of mode of data collection on estimates of multivariate models.

Two rival hypotheses can be developed about the effect of data collection mode on multivariate relationships between variables.

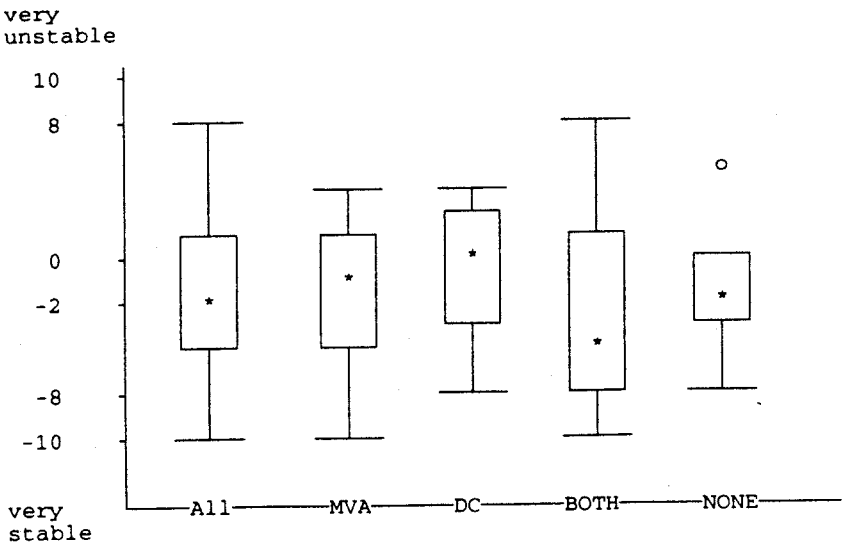
The first hypothesis asserts that even if mode effects may exist when univariate statistics are compared, this does not imply an effect on multivariate statistics. The reasoning is that the observed differences between the marginals of the univariate distributions just reflect a shift of position of a specific variable on the x- or y-axis, but that the shape of the bivariate distribution of any two variables (as reflected in the bivariate scatterplot) will not be altered. This is sometimes called the 'form-resistant correlation hypothesis' (cf. Krosnick/Alwin, 1987). This reasoning leads to the hypothesis that, if changes are detected in marginal distributions, multivariate statistics will remain comparatively stable.

The second hypothesis derives from statistical distribution theory, which states that, in general, higher order moments are less stable than first order moments. This implies that a few outliers in a specific sample can result in a dramatic change in statistics based on higher order moments, such as covariances and correlations. Thus the second hypothesis asserts that, if changes are detected in marginal distributions, multivariate statistics are expected to be even more unstable.

Which hypothesis is the most likely, remains to be seen. To get an indication of the common opinion on this matter, a small survey was conducted among experts in the fields of data collection methods multivariate analysis (members of german and dutch research groups). The general opinion was that multivariate structures were a bit more stable (the median of the 65 judgments was -2, on a scale from -10 = very stable to +10 = very unstable; the different expert groups did not differ significantly), but most experts had no strong opinion on the matter. Figure

1 below presents the distribution of the responses in a box plot.

Figure 1: Judgments by experts about stability of covariance structures



Experts: MVA = Multi Variate Analysis; DC = Data Collection; BOTH = both MVA + DC; NONE = neither MVA nor DC.

Stability judged on scale from -10 = very stable, 0 = don't know, 10 = very unstable.

* = median

o = outlier

In this paper we investigate the potential influence of data collection method on the parameter estimates of two substantive models: a model about loneliness and a model about subjective well-being. Two different aspects of structural modeling are investigated: the loneliness model is a causal model of the determinants of loneliness, the subjective well-being model is a factor analysis model of the structure of well-being.

2 Data Collection

In the autumn of 1989 a controlled field experiment was conducted. The data were collected using three data collection methods: a mail survey, a (paper-and pencil) telephone interview, and a face to face interview. Care was taken to implement each data collection method as optimal as possible. In the mail survey condition Dillman's TDM (Dillman, 1978; De Leeuw/Hox, 1988) was followed completely, including a third and last reminder by certified mail. Twenty specially trained interviewers conducted the interviews. Ten randomly assigned interviewers started with the telephone interviews and then conducted face to face interviews; the other ten started with face to face interviews. All respondents received an advance letter. The interviewers used a standardized script in asking for respondent cooperation. In both the telephone survey and the face to face survey condition the request for cooperation was made by telephone. At least seven call-backs were made.

The subject of the questionnaire was well-being. It included questions about general satisfaction, loneliness, and happiness. In addition, a large number of background questions were asked, including questions on income and job situation. The questionnaire contained the 11-item De Jong-Gierveld loneliness scale (De Jong-Gierveld/Kamphuis, 1985), a condensed eight-item form of Brinkman's self-evaluation scale (Brinkman, 1977; Dykstra, forthcoming), and a balanced extension of Bradburn's affect balance scale consisting of a nine-item positive affect scale and a nine-item negative affect scale (Bradburn, 1969; Hox, 1986).

A random stratified sample from the telephone directory of the Netherlands was used. This sample was randomly split in three parts. For

the complete sample (respondents and nonrespondents) detailed background information was available based on the Dutch zip code system. The Dutch zip codes form an extremely fine grid; as a consequence aggregated information was available about socioeconomic status (SES), income, type of household, and type of community for clusters of, on average 15 households each.

The mail survey resulted in a final response rate of 68%. The face to face interview had a response rate of 51%, and the telephone interview had a response rate of 66%. We investigated the possibility of selective nonresponse using the auxiliary zip code information. Respondents and non respondents did differ slightly in affluence. However, no interaction effects with mode of data collection were observed: although the response rates differ, characteristics of respondents and non-respondents do not differ across the modes.

When the respondents in the three conditions were compared on their answers to socio-demographic questions, the only statistically significant differences observed over modes concerned the variables gender and marital status. No significant differences were detected across modes for important variables such as age, education and previous interview experience (testing was done at the 5% level).

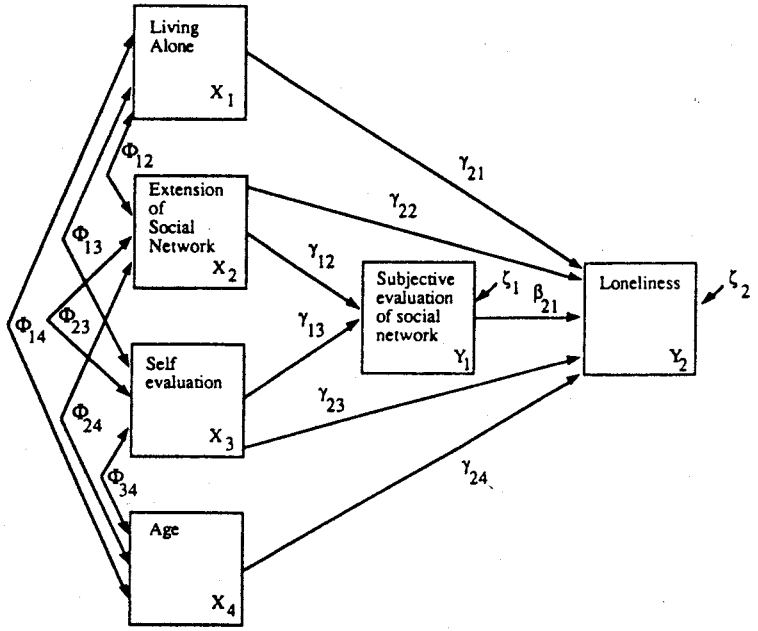
For more details on the data collection procedures and the nonresponse see De Leeuw (1992).

3 Method

Two different substantive models will be used to investigate the effect of data collection method on estimated relationships.

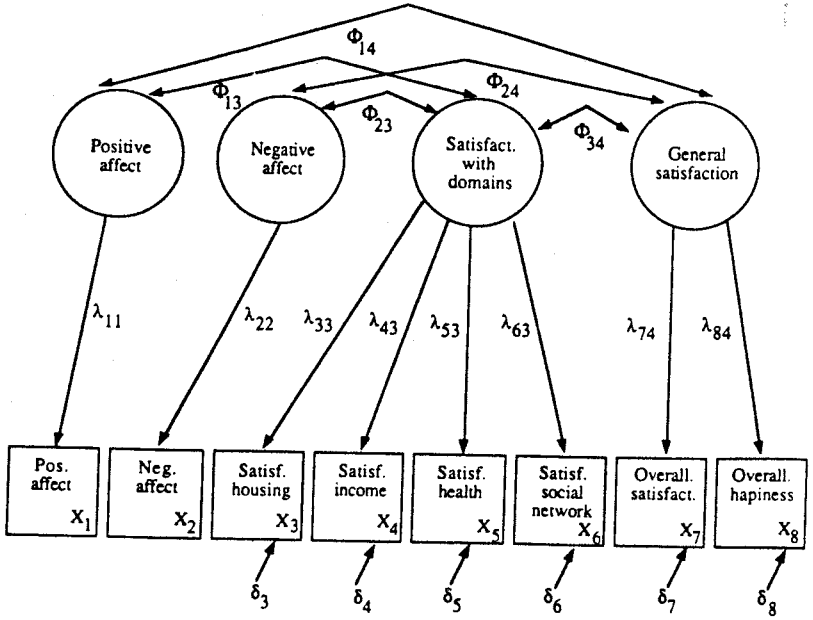
The first model - a causal structural equation model about the determinants of loneliness- is derived from De Jong-Gierveld (1987). In our model loneliness is negatively determined by the extension of the social network (number of important relationships), the amount of satisfaction with the social network, and a positive self-evaluation. Loneliness is (positively) determined by living alone and age (see Figure 2).

Figure 2: Loneliness Model



The second model – a measurement model on the structure of well-being – is derived from Burt et al. (Burt/Wiley/Minor/Murray, 1978; Burt/Fischer/Christman, 1979). Four dimensions are distinguished (see Figure 2). A general satisfaction dimension measured by two global variables, satisfaction with specific domains dimension measured by four variables on satisfaction with domains of life activity, a positive affect dimension perfectly measured by a (positive) affect variable, and a negative affect dimension perfectly measured by a (negative) affect variable. The positive and negative affect dimensions are uncorrelated (cf. Bradburn, 1969; Hox, 1986; see Figure 3).

Figure 3: Well-being Model



The following analysis strategy was used. First, for each model we examined whether the covariance matrices differed for the three data collection methods. This was followed by a series of multi-group analyses to investigate whether the models have the same parameter values for the mail survey, the telephone survey, and the face to face survey (Bollen, 1989, chap. 8; Jöreskog/Sörbom, 1989, chap. 9).

We started with the strictest model in which all parameters were assumed to be invariant over the three groups (i.e., the mail, the telephone, and the face to face survey).

The next model includes information about the reliability of measurement. Preliminary analyses had indicated that the reliability of the multi-item scales differed across the data collection methods: the mail survey showed the most reliable results, while the telephone survey was the least reliable (De Leeuw, 1991). Therefore, in the next model we allowed differences in measurement errors between the groups. For the multi-item scales the reliability estimates under the congeneric test model were used to estimate the variance of the corresponding measurement errors ϵ and δ (Bollen, 1989, p. 168). Furthermore, in the well-being model multiple observed variables were available for the latent variables 'general satisfaction' and 'satisfaction with specific domains'. For these observed variables, the estimated variances of the measurement errors Θ_δ were allowed to differ across groups.

In the next step, invariance restrictions between groups were only imposed on parameter estimates for the two interview modes (face to face and telephone). The model for the self-administered mail survey group was only restricted to have the same pattern as the two interview groups; but the loadings were allowed to differ in this group.

Finally, for all three groups the only restrictions concerned the form (same dimensions and patterns); all parameter estimates were allowed to differ in all groups.

To compare subsequent models the overall χ^2 and the overall root mean squared error was calculated. Furthermore the normed incremental fit index Δ was calculated (Bentler/Bonett, 1980). Δ measures the proportionate reduction in the χ^2 values when moving from a baseline model to the maintained model (Bollen, 1989, p. 270). As a baseline model the most restrictive model (model 1: all loadings invariant in all groups) is used. Furthermore, in most cases the subsequent models are nested within each other. For two nested models the difference in χ^2 is again χ^2 distributed with a df equal to the difference in df for the two models. This makes it possible to test if the improvement of fit is significant.

4 Results

4.1 The loneliness model

The loneliness model analyzed in this study is a causal model with four exogenous variables (living alone, number of important relationships, self-evaluation, and age) and two endogenous variables (evaluation of social network and loneliness). The loneliness model is a path model with observed variables. See also Figure 2 above.

For each data collection method (mail, telephone and face to face survey) a covariance matrix was computed. The covariance matrices were significantly different in the three groups (p=.00). Given this result, it is not surprising that the strictest model (model 1), which constrains all parameter estimates to be equal across all groups, did not fit.

In model 1 the measurement error variances were all fixed at zero. In the next model (model 2), estimates of the measurement error variance of the multi-item scales for self-esteem and loneliness were placed in the error-variance matrixes Θ_δ and Θ_ϵ; different values for each group were used based on the reliability estimates under the congeneric test model. This does not improve the fit of the model, therefore the subsequent models do not include the estimates of the measurement error.

The next step constrains all parameters to be invariant for the face to face and the telephone interview group only. In the mail survey group the parameter matrices are only constrained to have the same dimensions and patterns as in the two interview groups (model 3). This model has a reasonable fit (see Table 1). Since model 3 is nested in model 1 we can use the difference in χ^2 to test whether the increase in fit is statistically significant. Although the value of the incremental fit index is substantial (.39), the difference in χ^2 between model 1 and model 3 turns out to be not significant (p=.08).

In the final step (model 4), the restrictions are freed even further. In model 4 the only constraints are on the pattern of the parameter matrices. The same dimension and pattern is assumed, without restricting any of the nonfixed parameters to have the same value across groups.

Model 4 shows a good fit. Compared to model 1 the fit is significantly better (p=.02). Also, compared to model 3 the fit of model 4 is better (p=.04).

Table 1 summarizes the results of the consecutive modeling steps:

Table 1: Model Fit: Three Group Path Analysis Loneliness

Model	Restriction	χ^2	DF	P-VALUE	RMSR	Δ
(1)	Mail=FtF=Tel	39.8	24	.02	1.12	-
(2)	Mail=FtF=Tel/α	39.4	24	.03	1.06	.01
(3)	Mail$\approx$FtF=Tel	24.3	15	.06	1.10	.39
(4)	Mail$\approx$FtF$\approx$Tel	6.4	6	.38	.46	.84

(For each model the overall χ^2, degrees of freedom (DF), p-value, and the root mean squared residual (RMSR) is presented. Δ is the normed incremental fit index (evaluated against model 1, the strictest model). /α indicates error variances estimated from scale reliability.)

When comparing over groups, unstandardized parameter estimates are preferred (Bollen, 1989, p. 126). For the least restrictive model (model 4) the unstandardized parameter estimates are given in Table 2. To interpret the relative importance of the parameter estimates, it is then essential to know the scale on which the variables are measured. For loneliness the minimum score is 0 and the maximum score is 11, and the self-esteem score ranges from 0 to 8. The variable 'living alone' ranges from 1 (living with a family) to 3 (living completely alone). Extension of social network is a count of the number of important relations with a minimum of 0. Age is measured in years. Satisfaction with social network is measured on a five-point scale ranging from 1 (completely unsatisfied) to 5 (completely satisfied).

Table 2: Three Group Same Pattern Model (Mail$\approx$FtF$\approx$Tel) Loneliness

Parameter	Mail		Face to Face		Telephone	
β_{21}	-2.11	(.17)	-1.29	(.16)	-1.37	(.19)
γ_{21}	0.55	(.33)	0.51	(.30)	0.76	(.30)
γ_{22}	-0.29	(.10)	-0.30	(.11)	-0.23	(.12)
γ_{12}	0.08	(.04)	0.15	(.04)	0.05	(.04)
γ_{13}	0.09	(.03)	0.10	(.03)	0.05	(.03)
γ_{23}	-0.18	(.07)	-0.28	(.07)	-0.37	(.07)
γ_{24}	0.00	(.01)	0.03	(.01)	-0.00	(.01)
ψ_{11}	0.75	(.07)	0.83	(.08)	0.62	(.06)
ψ_{22}	4.58	(.44)	4.58	(.43)	5.33	(.48)
R^2_{y1}	0.08		0.11		0.02	
R^2_{y2}	0.52		0.41		0.29	

Unstandardized ML estimates for the mail, face to face, and telephone condition. Standard errors in parentheses.

When we inspect Table 2 we see that the major differences between data collection methods occur for the parameters β_{21}, γ_{12}, γ_{13}, γ_{23}, and γ_{24}.

To facilitate the interpretation, parameters are often standardized. Table 3 presents the same parameter estimates as Table 2, standardized to a common metric for the three groups (the parameter estimates are standardized on the pooled estimated variances under the fitted model, to preserve across groups comparability, cf. Jöreskog/Sörbom, 1989, p. 238).

Table 3: Three Group Same Pattern Model (Mail≈FtF≈Tel) Loneliness

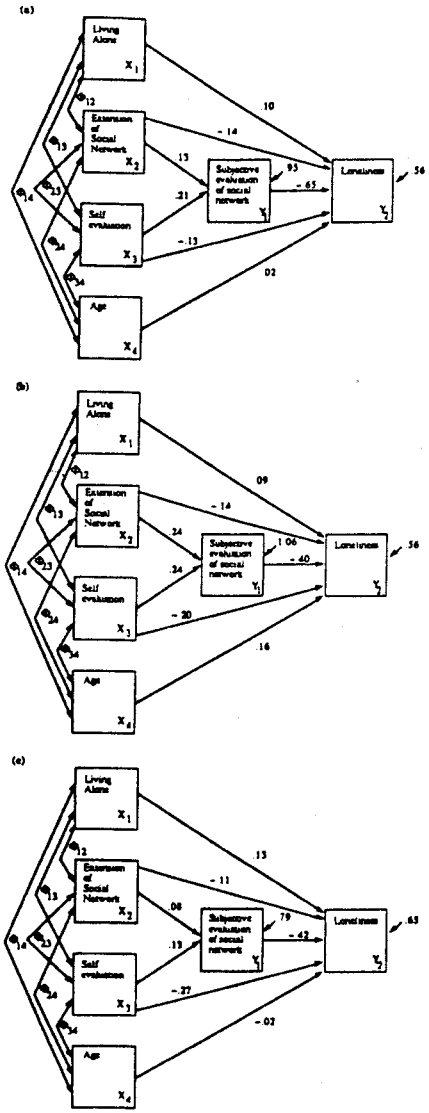
Parameter	Mail	Face to Face	Telephone
β_{21}	−0.65	−0.40	−0.42
γ_{21}	0.10	0.09	0.13
γ_{22}	−0.14	−0.14	−0.11
γ_{12}	0.13	0.24	0.08
γ_{13}	0.21	0.24	0.13
γ_{23}	−0.13	−0.20	−0.27
γ_{24}	0.02	0.16	−0.02
ψ_{11}	0.95	1.06	0.79
ψ_{22}	0.56	0.56	0.65

ML estimates standardized to a common metric.

Again, the most important differences between data collection methods occur for the parameters β_{21}, γ_{12}, γ_{13}, γ_{23}, and γ_{24}.

The differences shown in Table 3 are most likely large enough to influence the substantive interpretation of the results. Figure 4 shows the standardized parameter estimates for all groups in graphical form.

Figure 4: Standardized parameter estimates loneliness model (model 4) for (a) Mail Survey, (b) Face to Face Interview, (c) Telephone Interview



Again, the differences appear large enough to have a substantial influence on the substantive interpretation of the results.

4.2 The well-being model

The well-being model is a confirmative factor analysis mod[illegible] four factors (positive affect, negative affect, domain satisfaction, [illegible]eral satisfaction) measured by eight observed variables. See als[illegible]re 2. The original model, as published by Burt et al. (1978), is not identified; for a discussion of restrictions to make the model identified see Burt et al. (1979). In our version of the model, the variance of the factors is fixed at 1.00, and the measurement error variances of the two observed variables positive affect and negative affect are fixed.

For each data collection method (mail, telephone and face to face survey) a covariance matrix was computed. The covariance matrices were significantly different in the three groups (p=.00). Given this result, it is not surprising that the strictest model (model 1), which constrains all parameter estimates to be equal across all groups, did not fit.

In model 1 the measurement error variances for the two observed variables positive affect and negative affect were fixed at zero, all other measurement error variances were constrained to be equal across the three groups. In the next model (model 2), the measurement error variances of the observed variables for the factors 'domain satisfaction' and 'general satisfaction' were estimated separately in the three groups. This results in a model which fits much better than the first model, although the overall fit is still not good.

The next model (model 3) estimates the error variances for the two remaining observed variables (positive affect and negative affect) from the reliability estimates under the congeneric test model. This results in a slightly better fit.

In the next model (model 4) all parameters are constrained to be invariant for the face to face and the telephone interview group only. In the mail survey group the parameter matrices are only constrained to have the same dimensions and patterns as in the two interview groups. This

model fits better than model 2 and 3, which constrain the factor loadings and correlations, but allow the measurement errors to differ across all groups.

The next model (model 5) again allows the variances of the measurement errors (δ) of the observed variables for domain satisfaction and general satisfaction to vary across groups. Model 5 can be compared statistically with model 3, and fits significantly better than model 3 (p=.00). The next model (6) also estimates the fixed error variances of observed positive and negative affect using reliability estimates. This results in a slightly better fit than model 4.

In the final step (model 7), the restrictions are freed even further. In model 7 the only constraints are on the pattern of the parameter matrices. The same dimension and pattern are assumed, without restricting any of the nonfixed parameters to the same value across groups. Compared to model 2 (identical loadings and correlations, different measurement errors) the fit is significantly better (p=.00). Also, compared to model 4 (restrictions across face–to–face and telephone conditions) the fit of model 7 is better (p=.00). Compared to model 5 (restrictions across face–to–face and telephone conditions, different measurement errors) the fit of model 7 is also better (p=.03). However, the overall fit of model 7 is still not satisfactory.

For an overview of the fit statistics of the successive models see Table 4 below.

Table 4: Model Fit: Three Group Factor Analysis Well-Being

Model Restriction	χ^2	DF	P-Value	RMSR	Δ
(1) Mail=FtF=Tel	220.1	89	.00	.21	-
(2) Mail=FtF=Tel/δ	149.1	77	.00	.21	.32
(3) Mail=FtF=Tel/$\delta + \alpha$	148.6	77	.00	.21	.32
(4) Mail$\approx$FtF=Tel	131.1	70	.00	.14	.40
(5) Mail$\approx$FtF=Tel/δ	117.6	64	.00	.14	.47
(6) Mail$\approx$FtF=Tel/$\delta + \alpha$	117.2	64	.00	.13	.47
(7) Mail$\approx$FtF$\approx$Tel	93.0	51	.00	.10	.58

(For each model the overall χ^2, degrees of freedom (DF), p-value, and the root mean squared residual (RMSR) is presented. Δ is the normed incremental fit index (evaluated against the strictest model 1). /δ indicates different error variances; /α indicates error variances estimated from scale reliability.)

When comparing over groups, unstandardized parameter estimates are preferred (Bollen, 1989, p. 126). For the least restrictive model (model 7) the unstandardized parameter estimates are given in Table 5. To interpret the relative importance of the parameter estimates, it is important to know the scale on which the variables are measured. Positive and negative affect are measured by 9-item scales, with a range from 0 (lowest) to 9 (highest). The domain satisfactions and global satisfaction and are measured by single 5-point questions. Global happiness is measured by a single 7-point (ladder) question.

Table 5: Three Group Same Pattern Model (Mail$\approx$FtF$\approx$Tel) Well-Being

Parameter	Mail		Face to Face		Telephone	
λ_{11}	2.29	(.11)	2.01	(.10)	1.81	(.09)
λ_{22}	2.14	(.10)	2.25	(.11)	2.07	(.10)
λ_{33}	0.33	(.07)	0.23	(.07)	0.09	(.07)
λ_{43}	0.42	(.07)	0.28	(.08)	0.34	(.09)
λ_{53}	0.27	(.06)	0.27	(.08)	0.25	(.08)
λ_{63}	0.41	(.06)	0.65	(.10)	0.21	(.07)
λ_{74}	0.60	(.04)	0.54	(.06)	0.47	(.05)
λ_{84}	1.01	(.07)	0.83	(.10)	0.91	(.11)
Φ_{31}	0.56	(.09)	0.39	(.09)	0.35	(.15)
Φ_{32}	-0.62	(.09)	-0.41	(.09)	-0.40	(.15)
Φ_{41}	0.45	(.05)	0.39	(.07)	0.42	(.07)
Φ_{42}	-0.46	(.05)	-0.52	(.07)	-0.40	(.08)
Φ_{43}	1.13	(.09)	0.68	(.11)	1.21	(.25)
Θ_{δ_3}	0.92	(.09)	0.69	(.07)	0.95	(.09)
Θ_{δ_4}	0.88	(.09)	0.92	(.09)	0.83	(.09)
Θ_{δ_5}	0.69	(.06)	0.78	(.08)	0.91	(.09)
Θ_{δ_6}	0.64	(.06)	0.54	(.11)	0.54	(.05)
Θ_{δ_7}	0.12	(.02)	0.23	(.05)	0.28	(.04)
Θ_{δ_8}	0.53	(.08)	1.23	(.15)	1.22	(.16)

Unstandardized ML estimates for the mail, face to face, and telephone condition. Standard errors are given in parentheses.

Comparatively large differences between the groups are found for the loadings of 'housing' and 'health' (λ_{33} and λ_{63}) on the domain satisfaction factor, and for the correlations of the domain satisfaction factor (factor 3) with the other factors (Φ_{31}, Φ_{32}, Φ_{43}). (The latter even shows two values outside the permitted range, which again indicates that there are serious problems with the overall model.)

In the well-being model, the variances of the factors have been fixed at 1.00. To facilitate the interpretation of the factor loadings, the observed variables parameters are often also standardized. Table 6 presents the same factor loadings as Table 5, with the observed variables standardized to a common metric for the three groups (to preserve across groups

comparability; standardization is based on the pooled variance estimates for the observed variables under the fitted model, cf. Jöreskog/Sörbom, 1989).

Table 6: Three Group Same Pattern Model (Mail≈FtF≈Tel) Well-Being

Parameter	Mail	Face to Face	Telephone
λ_{11}	1.12	0.98	0.88
λ_{22}	0.99	1.04	0.96
λ_{33}	0.34	0.24	0.09
λ_{43}	0.42	0.28	0.34
λ_{53}	0.29	0.29	0.27
λ_{63}	0.47	0.74	0.23
λ_{74}	0.85	0.77	0.66
λ_{84}	0.75	0.62	0.67
Φ_{31}	0.56	0.39	0.35
Φ_{32}	-0.62	-0.41	-0.40
Φ_{41}	0.45	0.39	0.42
Φ_{42}	-0.46	-0.52	-0.40
Φ_{43}	1.13	0.68	1.21
Θ_{δ_3}	1.01	0.76	1.04
Θ_{δ_4}	0.88	0.92	0.83
Θ_{δ_5}	0.80	0.91	1.06
Θ_{δ_6}	0.81	0.69	0.69
Θ_{δ_7}	0.23	0.47	0.55
Θ_{δ_8}	0.29	0.67	0.67

ML estimates are standardized to a common metric.

In Table 6 the loadings λ_{33} and λ_{63} show relatively large differences across the three groups. Figure 5 presents the information in this table in graphical form:

Figure 5: Standardized parameter estimates well-being model (model 7) for (a) Mail Survey, (b) Face to Face Interview, (c) Telephone Interview

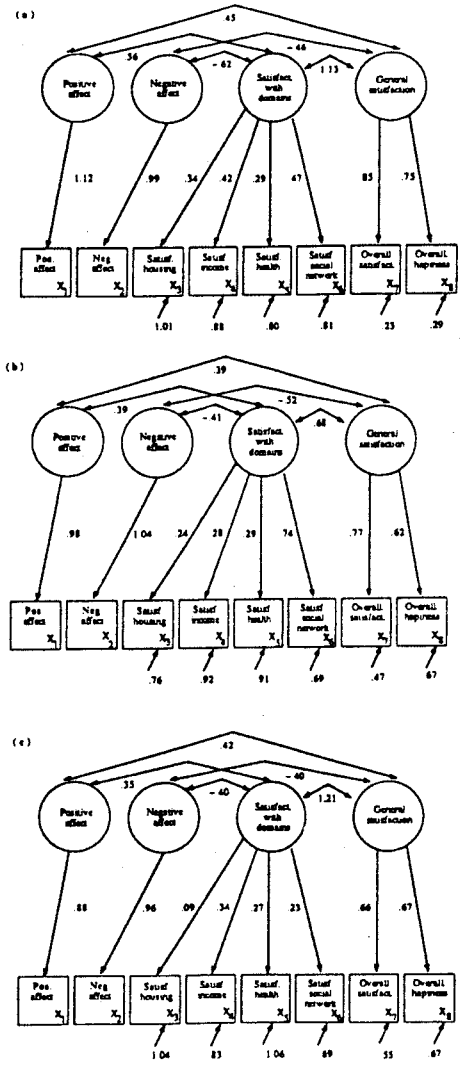


Figure 5 also shows that the differences between the data collection modes are not neglegible from a substantive point of view.

4.3 Explorative model search for the well-being model

The results for the well-being model show that even the least restrictive model does not fit well (p=.00). One strategy to search for a better fitting model is to employ an automatic model search procedure based on the modification index (Jöreskog/Sörbom, 1989). An interesting question is, how divergent the results of such an explorative procedure are across the three groups. To answer this question, an explorative model search was started with the least restrictive well-being model (model 7) as the start model. The model search was stopped when p-value of the difference between two successive models became larger than 1%. The outcome is a model which frees several factor loadings in the parameter matrix lambda. The results of the final model are presented in Table 7:

Table 7: Three Group Exploration (Mail≈FtF≈Tel) Well-Being

Parameter	Mail		Face to Face		Telephone	
λ_{11}	2.29	(.11)	2.01	(.10)	1.81	(.09)
λ_{22}	2.14	(.10)	2.25	(.11)	2.07	(.10)
λ_{33}	0.33	(.07)	0.32	(.08)	-0.52	(.64)
λ_{43}	0.42	(.07)	0.36	(.09)	0.25	(.14)
λ_{53}	0.27	(.06)	0.24	(.08)	-0.07	(.08)
λ_{63}	0.41	(.06)	0.52	(.09)	0.15	(.09)
λ_{74}	0.60	(.04)	0.53	(.05)	0.47	(.05)
λ_{84}	1.01	(.07)	1.49	(.42)	0.93	(.11)
λ_{61}	0.19	(.07)	0.23	(.08)	-	
λ_{83}	-		-0.83	(.44)	-	
λ_{34}	-		-		0.84	(.61)
λ_{54}	-		-		0.38	(.12)
Φ_{31}	0.41	(.09	0.12	(.12)	0.49	(.23)
Φ_{32}	-0.66	(.09)	-0.43	(.09)	-0.50	(.25)
Φ_{41}	0.45	(.05)	0.35	(.07)	0.41	(.07)
Φ_{42}	-0.46	(.05)	-0.54	(.06)	-0.39	(.07)
Φ_{43}	1.13	(.09)	0.77	(.11)	1.56	(.77)
Θ_{δ_3}	0.92	(.09)	0.64	(.07)	1.34	(.39)
Θ_{δ_4}	0.88	(.09)	0.87	(.09)	0.88	(.10)
Θ_{δ_5}	0.68	(.06)	0.79	(.08)	0.90	(.09)
Θ_{δ_6}	0.62	(.06)	0.60	(.08)	0.56	(.06)
Θ_{δ_7}	0.12	(.02)	0.25	(.04)	0.27	(.04)
Θ_{δ_8}	0.53	(.08)	0.92	(.26)	1.18	(.16)

Unstandardized ML estimates for the mail, face to face, and telephone condition. Standard errors are given in parentheses.

Table 7 shows several problems (parameter estimates outside permitted boundaries), which should be fixed in further model runs. Even without such additional fix-ups it is obvious that the results diverge considerably across the three modes.

5 Summary and discussion

To investigate the potential influence of data collection method on the estimates of structural relationships between variables we compared two substantive structural-equation models across three different data collection methods: a loneliness model and a well-being model. The loneliness model is a causal model with four exogenous variables (living alone, extension of social network, self-evaluation, and age) and two endogenous variables (evaluation of social network and loneliness). It is a path model with observed variables. The well-being model is a confirmatory factor analysis model with four factors (positive affect, negative affect, domain satisfaction, and general satisfaction), measured by eight observed variables.

Two rival hypotheses were investigated. The first hypothesis states that, although small mode effects are in general found for the univariate distributions of observed variables, the multivariate estimates will remain stable (form resistant correlation hypothesis). The second hypothesis states that if (small) mode effects are found in univariate distributions, multivariate estimates will show even larger effects (instability of higher order moments hypothesis).

A small survey among experts in the field of data collection and experts in the field of multivariate analysis shows a small preference for the form resistant correlation hypothesis.

The results of a LISREL multi-group analysis lend support to the other hypothesis, the instability of higher moments hypothesis. For both the loneliness-model and the well-being model the strictest statistical model was rejected; this model assumes invariance of all parameters over the three groups (i.e., the mail, the telephone, and the face to face survey). A less strict model was more appropriate. This model assumes the same dimension and pattern across groups without restricting any of the non-fixed parameters. Comparison of the estimates under this model for the two substantive models gives cause for some concern about the 'robustness' of substantive interpretations of models based on data collected by different modes.

For the loneliness model, the least restrictive (same pattern) model had a good statistical fit. The loneliness model is a path-model in which the loneliness scale is the major dependent variable. In both the mail survey and the face to face interview group the proportion of variance explained was relatively high (.52 and .41), in the telephone condition this figure was only lower at 0.29. Thus, the same set of explanatory variables explains far less variance in the telephone survey condition. Also, the relative importance of individual explanatory variables varies considerably across data collection methods (cf. Figure 4). In the mail survey condition the influence of subjective evaluation of the social network on feelings of loneliness is considerable larger than in either the face to face or the telephone condition. However, in all three groups the subjective evaluation of the social network is the most important determinant of feelings of loneliness. A striking difference is found when we consider the variable age. Only in the face to face condition is age a relative important determinant of feelings of loneliness.

The well-being model (a factor model with four factors) showed a less satisfactory overall statistical fit for the least restrictive (same pattern) model specification (p=.00). However, the value of the root mean squared residual (.10) and the relative size of the χ^2 and the degrees of freedom (χ^2/df=1.82) suggest that this least restrictive model is still acceptable.

When we consider the parameter estimates under this model we see again a marked difference in the relative importance of the variables. In the mail survey condition the observed variable (satisfaction with) 'social network' is the most important variable for the dimension 'domain satisfaction (λ = .47)', immediately followed by 'income'. 'Housing' and 'health' are less important. In the face to face interview-condition the most important variable is 'social network'(λ = .73); the variables 'health', 'income', and 'housing' hardly differ in relative importance. In the telephone condition 'income' is the most important variable for the dimension domain satisfaction (.34), while 'social network' is the third important variable (.24). See also figure 5, which shows the parameter estimates standardized to a common metric for the three groups.

As mentioned above the statistical fit for even the least restrictive (same

pattern) model was not quite satisfactory. Exploratory analyses in which restrictions between groups were automatically freed based on their modification indices resulted in a statistically acceptable model. In this model the structure of well-being diverges even more across groups, because different factor loadings λ are freed.

6 References

Baumgartner, R. M.; T. A. Heberlein (1984): Recent Research on Mailed Questionnaire Response Rate. D. C. Lockhardt (Ed.): Making Effective Use of Mailed Questionnaires. San Francisco.

Bentler, P. M.; D. G. Bonett (1980): Significance Tests and Goodness-Of-Fit in the Analysis of Covariance Structures. Psychological Bulletin. 88. 588-600.

Bollen, K. E. (1989): Structural Equations with Latent Variables. New York.

Bradburn, N. M. (1969): The Structure of Well-Being. Chicago.

Brinkman, W. (1987): Een assertiviteitsschaal II [Measuring Assertivity II]. Amsterdam.

Burt, R. S.; J. A. Wiley; M. J. Minor; J. R. Murray (1978): Structure of Well-Being. Form, Content, and Stability over Time. Sociological Methods and Research. 6. 365-407.

Burt, R. S.; M. G. Fischer; K. P. Christman (1979): Structures of Well-Being. Sufficient Conditions for Identification as Restricted Covariance Models. Sociological Methods and Research. 8. 111-120.

De Jong-Gierveld, J. (1987): Developing and Testing a Model of Loneliness. Journal of Personality and Social Psychology. 53. 119-128.

De Jong-Gierveld, J.; F. Kamphuis (1985): The Development of a Rasch-Type Loneliness Scale. Applied Psychological Measurement. 9. 289-299.

De Leeuw, E. D. (1991): The Influence of Data Collection Procedure on Psychometric Reliability and Scaling Properties. (Response Effects in Surveys, Technical Report No 5): Amsterdam.

De Leeuw, E. D. (1992): Data Quality in Mail, Telephone, and Face to Face Surveys. Ph.D. Thesis. Amsterdam.

De Leeuw, E. D.; J. J. Hox (1988): The Effects of Response-Stimulating Factors on Response Rates and Data Quality in Mail Surveys; A Test of Dillman's Total Design Method. Journal of Official Statistics. 4. 241-250.

De Leeuw, E. D.; J. J. Hox; J. Van der Zouwen (1989): Data Quality in Mail Surveys as Compared to Face to Face and Telephone Interviews: A Meta Analysis of the Research Literature. Kwantitatieve Methoden. 31. 123-146.

De Leeuw, E. D.; J. Van der Zouwen (1988): Data Quality in Telephone and Face to Face Surveys. R. M. Groves et al. (Eds.): Telephone Survey Methodology. New York. 283-299.

Dillman, D. A. (1978): Mail and Telephone Surveys. The Total Design Method. New York.

Dillman, D. A. (1991): The Design and Administration of Mail Surveys. Annual Review of Sociology. 17. 225-249.

Dillman, D. A.; J. Tarnai (1988): Administrative Issues in Mixed Mode Surveys. R. M. Groves et al. (Eds.): Telephone Survey Methodology. New York. 509-528.

Dykstra, P. A. (forthcoming): Alternative for the Absence of a Partner: The Presence of Supportive Relationships and the Desire for Independence as Factors that Serve to Mitigate Loneliness. Ageing and Society.

Goyder, J. (1985): Face to Face Interviews and Mailed Questionnaires: The Net Difference in Response Rate. Public Opinion Quarterly. 49. 234-252.

Groves, R. M. (1989): Survey Errors and Survey Costs. New York.

Groves, R. M.; R. L. Kahn (1979): Surveys by Telephone. A National Comparison with Personal Interviews. New York.

Hox, J. J. (1986): Het gebruik van hulptheoriën bij operationaliseren [Using Auxiliary Theories for Operationalization. A Study of the Construct of Subjective Well-Being]. (Doctoral Dissertation). Amsterdam.

Jöreskog, K. G.; D. Sörbom (1989): LISREL 7. A Guide to the Program and Applications (Second Edition). Chicago.

Krosnick, J. A.; D. F. Alwin (1987): A Cognitive Theory of Response-Order Effects. Public Opinion Quarterly. 51. 201–219.

Lyberg, L.; D. Kasprzyk (1991): Data Collection Methods and Measurement Error: An overview. P. P. Biemer et al. (Eds.): Measurement Errors in Surveys. New York. 237–257.