A NEW APPROACH TO EVALUATING THE QUALITY OF MEASUREMENT INSTRUMENTS: THE SPLIT-BALLOT MTMM DESIGN

*Willem E. Saris**

Albert Satorra[†]

Germà Coenders[‡]

Two distinctly different quantitative approaches are used to evaluate measurement instruments: the split-ballot experiment and the multitrait-multimethod (MTMM) approach. The first approach is typically used to indicate whether variation in the method causes differences in the response distribution; the second approach evaluates the reliability and validity of different methods. The new approach, suggested in this paper, combines the more attractive features of both methods. The strength of the split-ballot experiment is its use of independent random samples from the same population to provide information about differences in response distributions. This is also possible with the new approach, but this approach provides more detailed information about the reasons

We thank the anonymous reviewers for their comments and the support of the Spanish Ministry of Science and Technology through the grant SEC2003-04476. Direct correspondence to Willem Saris, ESADE, Universitat Ramon Llull, Avenue de Pedralbes 60-62, 08034 Barcelona, Spain, e-mail: wsaris@planet.nl.

*Universitat Ramon Llull, Barcelona, and University of Amsterdam, The Netherlands

[†]Universitat Pompeu Fabra, Barcelona,

[‡]Universitat de Girona, Spain

for the differences. The MTMM approach provides information about reliability and validity on the basis of repeated observation of the same traits using different methods. This information is also provided by the new design. The difference is that the new approach reduces the need for repeated observations of the same trait. Each sample is provided with a different combination of only two methods and the complete model with all methods is estimated as a multiple-group model. This reduces the burden for respondents and also reduces memory and order effects. Alternative designs and estimation methods are discussed, their efficiency is analyzed, and illustrations are provided.

In 1959 Campbell and Fiske suggested the multitrait-multimethod (MTMM) design for evaluating the validity of measurement instruments. At first, the correlations were interpreted directly, as suggested by the authors, but soon structural equation models were developed for evaluation of measurement instruments. A review of all these models can be found in Wothke (1996). Among them is the *confirmatory factor analysis* model for MTMM data (Althausen, Herberlein, and Scott 1971; Alwin 1974; Werts and Linn 1970). An alternative parameterization of this model proposed by Saris and Andrews (1991) is known as the *true score* (TS) model, while the *correlated uniqueness* model was put forward by Kenny (1976), Marsh (1989), and Marsh and Bailey (1991). Rather different models with what are called multiplicative method effects were suggested by Campbell and O'Connell (1967), Browne (1984), and Cudeck (1988). Coenders and Saris (1998, 2000) showed that the multiplicative model can be formulated as a special case of the correlated uniqueness model of Marsh (1989).

Although the MTMM approach is accepted as a useful tool and is widely used, much attention has been given to its frequent problems of nonconvergence, underidentification, or improper solutions for the confirmatory factor analysis model (Andrews 1984; Bagozzi and Yi 1991; Brannick and Spector 1990; Kenny and Kashy 1992; Marsh and Bailey 1991; Saris 1990). Grayson and Marsh (1994) showed that confirmatory factor analysis models with correlated method factors are usually underidentified, which may explain why these problems occur. Eid (2000) discussed these problems again and suggested an alternative model with one factor fewer than usual. Conversely, models with correlated traits and uncorrelated methods (CTUM), which

should not have the same problem, exist. This solution was also suggested by Andrews (1984) and Saris (1990). A recent study confirmed that a model equivalent to the CTUM model does indeed suffer from few problems (Corten et al. 2002).

A more severe drawback of the standard MTMM approach is that at least three methods must be included to prevent even more severe problems of empirical underidentification (Kenny 1976; Scherpenzeel 1995), meaning that the same respondents have to be asked about the same trait three times. Van Meurs and Saris (1990) showed that respondents do not remember their previous answers if at least 20 minutes elapse between consecutive measures, provided that questions with similar format are asked in the interim and that the opinions of respondents are not extreme. If this rule is applied to MTMM designs, over 40 minutes of interview time is required for cross-sectional studies. Even if memory effects are ruled out by a generous spacing of questions, there remains the problem that the response burden for the respondents is quite high. Therefore, the second and third measurements may not be as accurate as the first, merely due to the order in the questionnaire.

We believe this problem of two repeated measures threatens the MTMM approach more seriously than the technical problems of nonconvergence and improper solutions. Therefore, we suggest new designs for MTMM studies that reduce the response burden by means of using different combinations of only two methods in multiple groups and estimating the MTMM model under the multiple-group structural equation modeling (SEM) approach. The use of multiple groups brings the design close to the popular split-ballot designs introduced by survey researchers in the first half of the last century (for an overview, see Schuman and Presser 1981).

We will show that our new design combines the benefits of the split-ballot approach, providing information on differences in distributions for different forms, and the MTMM approach. It enables researchers to evaluate measurement reliability and validity, and does so while reducing the response burden. Section 1 explains the classic MTMM design. Section 2 discusses various alternative designs and the estimation and testing of the models. Section 3 introduces two empirical examples that illustrate the methods proposed. Section 4 covers the identification and efficiency of the designs discussed. The paper concludes with a discussion.

TABLE 1
The Correlations Between the Nine Variables of the MTMM Experiment with
Respect to Satisfaction with Political Outcomes

	Form 1			Form 2			Form 3		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
Form 1									
Q1	1.00								
Q2	.481	1.00							
Q3	.373	.552	1.00						
Form 2									
Q1	-.626	-.422	-.410	1.00					
Q2	-.429	-.663	-.532	.642	1.00				
Q3	-.453	-.495	-.669	.612	.693	1.00			
Form 3									
Q1	-.502	-.347	-.332	.584	.436	.438	1.00		
Q2	-.370	-.608	-.399	.429	.653	.466	.556	1.00	
Q3	-.336	-.406	-.566	.406	.471	.638	.514	.558	1.00
Means	2.42	2.71	2.45	5.26	4.37	5.13	2.01	1.75	2.01
Standard dev.	.77	.76	.84	2.29	2.37	2.44	.72	.71	.77

1.1. A Possible Explanation

Given that the same people answer all questions, one explanation given for the differences between these correlations is measurement error. It is supposed that each method-trait combination has its own random errors and systematic errors, the latter called the *method effect*. Formally, this was specified by Saris and Andrews (1991) as

$$Y_{ij} = r_{ij}T_{ij} + e_{ij} \qquad \text{for } i = 1 - 3 \text{ and } j = 1 - 3 \qquad (1)$$

$$T_{ij} = v_{ij}F_i + m_{ij}M_j \qquad \text{for } i = 1 - 3 \text{ and } j = 1 - 3, \qquad (2)$$

where

- Y_{ij} is the measured variable (trait i measured by method j).
- T_{ij} is the stable component of the response Y_{ij} (also called the “true score”).
- F_i is the trait factor.
- M_j is the method factor, whose variance represents systematic method effects common for all traits but varying across individuals.

- e_{ij} is the random measurement error term for Y_{ij} , with zero mean and uncorrelated with other error terms, with method factors and with trait factors.
- The r_{ij} coefficients standardized can be interpreted as reliability coefficients (square root of test-retest reliability).
- When standardized, the m_{ij} coefficients represent method effects, while m_{ij}^2 equals, the explained variance in the true score by the method factor.
- The v_{ij} coefficients standardized are validity coefficients (with v_{ij}^2 representing the validity of the measure). Note that in this model the validity is reduced only by method effects since $v_{ij}^2 = 1 - m_{ij}^2$.

Figure 2 shows the same model for two traits measured with the same method and on the assumption that the measurement errors are independent of each other and independent too of the true scores and trait variables.

Path analysis can be used to show that the correlation between the observed variables $r(Y_{1j}, Y_{2j})$ is equal to the correlation of the variables we want to measure, F_1 and F_2 , multiplied by the reliability and validity coefficients of the two observed variables, plus the correlation due to the method effects multiplied by the reliability coefficients of the two measurements:

$$r(Y_{1j}, Y_{2j}) = r_{1j}v_{1j}r(F_1, F_2)v_{2j}r_{2j} + r_{1j}m_{1j}m_{2j}r_{2j}. \quad (3)$$

FIGURE 2. The measurement model for two traits measured with the same method.

The reliability and validity coefficients (r_{ij} and v_{ij}) are always smaller than 1; so, the lower the reliability, the larger the difference between the observed correlation and the correlation between the latent traits will be. Since the second term is typically positive, the method effects will usually inflate the correlation observed. This result suggests that the relatively low correlations observed for Form 1 and 3 in Table 1 are due to relatively high reliability of Form 2. However, Form 2 correlations also may be higher due to higher systematic method effects. Unless these data quality indicators are estimated, the reasons for the differences cannot be known.

1.2. *The Classic MTMM Design for Estimation of Reliability and Validity*

Clearly these coefficients cannot be estimated if only one measurement of each trait is available, since in this case there would be only one observed correlation available to estimate seven free parameters.

This is why an MTMM design with three traits, each measured with three different methods, as indicated above, was suggested. The 9×9 correlation matrix obtained is sufficient for calculation of all parameters. If this standard design is used, equation (2) represents the basic equation of the MTMM model and generates the following factor loadings structure for the standard MTMM design:

	F_1	F_2	F_3	M_1	M_2	M_3	
T_{11}	v_{11}			m_{11}			
T_{21}		v_{21}		m_{21}			
T_{31}			v_{31}	m_{31}			
T_{12}	v_{12}				m_{12}		
T_{22}		v_{22}			m_{22}		
T_{32}			v_{32}		m_{32}		
T_{13}	v_{13}					m_{13}	
T_{23}		v_{23}				m_{23}	
T_{33}			v_{33}			m_{33}	

(4a)

The specification of the structure of this matrix of loadings is commonly accepted, but Scherpenzeel and Saris (1997) suggest specifying additionally that

$$m_{ij} = m_m \quad \text{for all } i \quad (4b)$$

$$\text{var}(M_j) = 1 \quad \text{for all } j \quad (4c)$$

The consequence of these restrictions is that the method effects are the same for different traits measured by the same method. Many researchers do not introduce these restrictions.

Zero correlations between factors and the error terms is commonly accepted, but there is disagreement about the specification of the correlations between the different factors. Some authors leave all correlations free but mention that this leads to many problems (Kenny and Kashy 1992; Marsh and Bailey 1991). Others, like Andrews (1984) and Saris (1990), suggest that the trait factors should be allowed to correlate with each other, but their correlation with method factors should be restricted to zero, while the method factors should also be uncorrelated with each other. Using the latter specification, combined with the assumptions in (4b) and (4c), hardly any problems occur in practice (see Corten et al. [2002], who reanalyzed 79 MTMM experiments).

The specification of the model presented in equations (1) through (4) and standard ML estimation¹ (based on the covariance matrix) gave the results in Table 2 after standardization of latent and observed variables.

These results indicate that the second form of the questions has higher reliability coefficients than the other forms, with the method effect for this form being in between the two others.

¹This estimator is only the ML estimator if the distributional assumptions are satisfied. As long as we do not know whether that is the case, we cannot be sure that the estimates have the qualities of the ML estimators except under certain conditions specified by Arminger and Sobel (1990) and Satorra (1992, 2001). One assumption that is certainly not satisfied is the assumption of continuous variables. There has been a long debate, to which we have contributed (Coenders and Saris 1995; Coenders, Satorra, and Saris 1997; Saris, Van Wijk, and Scherpenzeel 1998), on whether such data should be analyzed on the basis of Pearson correlations or alternatives such as polychoric correlations (e.g., Olsson 1979; Jöreskog 1990). Our conclusion is that both are necessary since researchers use both kinds of summary measurements. If data are analyzed using Pearson correlations, the data quality corrections should also be based on these correlations. However, if polychoric correlations are used in the substantive research, the data quality corrections should also be based on analysis of polychoric correlations. This point has been made in Saris, Van Wijk, and Scherpenzeel (1998), which debates both analyses.

TABLE 2
The Standardized Estimates of the Parameters of the MTMM Model Specified
for the ESS Data of Figure 1.

	Validity Coefficients			Method Effects			Reliability Coefficients
	<i>F</i> ₁	<i>F</i> ₂	<i>F</i> ₃	<i>M</i> ₁	<i>M</i> ₂	<i>M</i> ₃	
<i>T</i> ₁₁	.93			.36			.79
<i>T</i> ₂₁		.94		.35			.85
<i>T</i> ₃₂			.95	.33			.81
<i>T</i> ₁₂	.91				.41		.91
<i>T</i> ₂₂		.92			.39		.94
<i>T</i> ₃₂			.93		.38		.93
<i>T</i> ₁₃	.85					.52	.82
<i>T</i> ₂₃		.87				.50	.87
<i>T</i> ₃₃			.88			.48	.84

Note: Chi-square = 31.0; d.f. = 21; *n* = 428.

Since the correlation between the first two traits was estimated to be .69, by using equation (3) we can easily verify that the measurement quality indicators can produce such different correlations as .48 for the first form, .64 for the second form, and .56 for the third form. This means that the observed differences of correlations are explained fully by differences in data quality between the different measurement procedures.²

In this case we used the True Score (TS) MTMM model specified by Saris and Andrews (1991). Many other models are discussed in the literature (Wothke 1996; Coenders and Saris 2000). The classic MTMM model is equivalent to the TS model (the difference lies only in the parameterization). More details of the relation between the TS model and the classic MTMM models are given in Appendix A.

2. ALTERNATIVE DESIGNS

Although the MTMM approach looks attractive, a major problem of this design is that the respondents have to answer questions about

²The same check could be done for the other correlations, which were respectively .66 for traits 1 and 3 and .74 for traits 2 and 3.

substantially the same questions three times. This might lead to a loss in precision because the respondents get annoyed or to greater precision because they had more time to think, or it might just induce correlated errors due to memory effects. Looking back to the correlation matrix in Table 1, we see that the correlations between the three variables are higher for the second and third methods than for the first method. So we might wonder if this is due to difference in method, as we have argued above, or to people having more time to think and realize that there are relationships between the questions, which would be an “occasion effect.” That the correlations for the third method are lower than for the second method could also be a “fatigue effect.”

There are two ways of coping with this problem. The first is to try to reduce the number of repeated observations. In this paper we concentrate on this approach. The other approach is to estimate the effect of the question order and try to correct for it. Scherpenzeel and Saris (1997) tackled these problems with a two-wave panel MTMM design, in which there were only two observations of the same trait in each wave and the order of the questions was changed randomly for the different respondents. One advantage of this design is that we can estimate the effect of the different occasions. Another advantage is that the response burden in each wave is reduced. The disadvantages are that the total response burden is increased by one extra measurement and that a frequently observed panel is required for the design. Although the design has been used in a large number of studies, thanks to the presence of a frequently observed panel, we think that this is not a solution that can be generally recommended.

Therefore, we suggest several other designs that could be used as alternatives. These designs reduce the number of observations per person but compensate for the “missing data by design” by collecting data from different subsamples of the population. This makes the designs look very similar to the frequently used split-ballot experiments, and therefore we have called the approach the split-ballot MTMM design (SB-MTMM).

2.1. *The Split-Ballot MTMM Design*

In the commonly used split-ballot experiments, random samples from the same population are given different versions of the same questions—i.e., each group gets one method. The split-ballot design makes it

possible to compare the response distributions for the different questions across forms of the question and hence to assess the relative bias (see Schuman and Presser 1981; Billiet, Loosveldt, and Waterplas 1986).

The SB-MTMM design also employs various random samples of the same population, but in each of the samples two forms of a question are used, which is one less than in the classic MTMM design and one more than in the commonly used split-ballot designs. This design, suggested by Saris (1998), combines the benefits of the split-ballot and MTMM approaches in that it enables researchers to evaluate measurement bias, reliability, and validity simultaneously, and it reduces response burden. A suggestion to use such split-ballot designs for structural equation models can be traced back to Arminger and Sobel (1990). A recent alternative, a more complex design in practical terms, has been suggested by Bunting, Adamson, and Mulhall (2002).

2.1.1. *The Two-Group Design*

The two-group SB-MTMM design is structured as follows. The sample is split randomly into two groups. One group has to answer three questions using form 1 (Method 1), while the other group is given the same questions but using form 2 (Method 2). In the last part of the questionnaire all respondents are again presented with the three questions, but now using form 3 (Method 3). The design can be summarized as follows:

	Time 1	Time 2
Sample 1	Form 1	Form 3
Sample 2	Form 2	Form 3

Thus under the two-group design, the researcher has to draw two comparable random samples from the same population and twice ask three questions about three traits in each sample. At Time 1, the two groups get a different form (method) of the three questions. At Time 2, after sufficient time has elapsed, the two groups get the same form of the three questions. Van Meurs and Saris (1990) suggested that 20 minutes of similar questions are enough to obtain independent measurements—i.e., where memory effects are negligible.

The questions at Time 1 match the design of the standard split-ballot design and therefore provide the same information about differences in response distributions between methods. Combined with the information at Time 2, this design can provide information on

TABLE 3
Samples Providing Data for Covariance Estimation

	Form 1	Form 2	Form 3
Form 1	Sample 1		
Form 2	None	Sample 2	
Form 3	Sample 1	Sample 2	Samples 1 and 2

reliability and validity and on method effects, while each respondent answers only two questions about the same trait, not three as was required in the classic MTMM design. This result is not immediately clear because the necessary information for the 9×9 covariance matrix comes from different groups and is incomplete by design, as can be seen in Table 3. The table shows the groups that provide data for estimating variances and correlations between questions using either the same or different forms.

In this case, unlike in the classic design, no covariances are obtained for Form 1 and Form 2 questions. These covariances are missing by design. Otherwise, all cells of the 9×9 matrix would be estimated on the basis of one or two samples, but different parts come from different samples. This design was proposed for the first time by Saris (1998) and his data have been reanalyzed by Saris and Krosnick (forthcoming).

It should be clear that each respondent is given the same questions only twice, reducing the response burden considerably. In large surveys we can even split the sample into more subsamples and in this way evaluate more than one set of questions. However, the covariances between Form 1 and Form 2 cannot be estimated, which results in a loss of degrees of freedom when estimating the model using this incomplete covariance matrix. This might make the estimation less efficient than in the standard design or in an approach where all covariances can be obtained, like the three-group design.

2.1.2. *The Three-Group Design*

The three-group design is like the previous design, except that three groups or samples are used instead of two. This leads to the following:

	Time 1	Time 2
Sample 1	Form 1	Form 2
Sample 2	Form 2	Form 3
Sample 3	Form 3	Form 1

Using this design, all forms of the questions are treated equally: All are measured once at the first and once at the second point in time. There are also no missing covariances in the covariance matrix, as can be seen in Table 4.

The major advantage of this approach is that all covariances can be estimated. Another advantage is that the order effects are cancelled out because each measurement occurs once at the first and once at the second position.

The major disadvantage of this approach is of a more practical nature. In this design the main questionnaire has to be prepared in three different forms for the three different groups. In addition, no method is used for all respondents, and thus comparable data cannot be produced for all respondents. This can be seen as a serious problem in the analysis because it reduces the sample size with respect to the relationships with other variables.³ This design was used for the first time by Kogovsek et al. (2002).

2.1.3. *Other SB-MTMM Designs*

Other designs can also be formulated along the principles indicated above. In principle, the effects of many different factors can be studied simultaneously, which also allows for the estimation of

TABLE 4
Samples Providing Data for Covariance Estimation

	Form 1	Form 2	Form 3
Form 1	Samples 1 and 3		
Form 2	Sample 1	Samples 1 and 2	
Form 3	Sample 3	Sample 2	Samples 2 and 3

³A possible alternative would be to add to the study a relatively small subsample. With the whole sample, we would use Method 1, the method expected to give the best results, in the main questionnaire. In a supplementary methodological questionnaire, Method 2 is used in one subgroup and Method 3 in another subgroup of the sample. In the extra subsample, we would use Method 2 for the main questionnaire and Method 3 in the methodological part. Thus Method 1 is available for everyone; as are all three combinations of the forms. In this way we could get an estimate of the complete covariance matrix for the MTMM analysis without harming the substantive analysis. However, this design is more expensive because of the additional subsample. The size of the subsamples is a matter for further research.

interaction effects. However, an alternative to such studies is the use of meta analysis of many separate MTMM experiments under different conditions. This approach was suggested by Andrews (1984), was further explored by Saris and Münnich (1995), and was applied by Rodgers, Andrews, and Herzog (1992), Költringer (1995), Scherpenzeel (1995), and Scherpenzeel and Saris (1997).

There is, however, one other design that deserves special attention. This SB-MTMM design makes use of exact replications of methods. Thus the occasion effects can also be studied without putting an extra response burden on the respondents. A possible design might be as follows:

	Time 1	Time 2
Sample 1	Form 1	Form 1
Sample 2	Form 1	Form 2
Sample 3	Form 2	Form 1
Sample 4	Form 2	Form 2

This is a complete four-group design for two methods and replications. It can be shown that this design can be reduced to an incomplete three-group design by leaving out either Sample 2 or 3 or alternatively Sample 1 or 4. With these incomplete SB-MTMM designs, all parameters of the standard MTMM design can also be estimated. The attractiveness of this design is that we can even estimate the specific variation of occasion, which is not possible in the previous two designs. It is possible only if exact repetition of the same measurements is included in the design. To estimate these effects, we have to extend the model specified in (1) and (2) by an occasion-specific component as

$$Y_{ijk} = r_{ijk}T_{ijk} + e_{ijk} \quad \text{for } i = 1 - 3, j = 1 - 3 \text{ and } k = 1 - 2 \quad (5)$$

$$T_{ijk} = v_{ijk}F_i + m_{ijk}M_j + o_{ijk}O_k \quad \text{for } i = 1 - 3, j = 1 - 3 \text{ and } k = 1 - 2, \quad (6)$$

where o_{ijk} represents the effect of the k th occasion specific factor, and O_k represents the specific factor for the k th occasion.

Clearly, a design including three different methods can be developed in a similar way. However, further discussion of this possibility here would lead too far. We hope that it is also clear that the major advantage of these designs is the reduction of the response

burden from three to two observations, which is important in practice. To show that these designs can be used, we now discuss the estimation of the parameters on the basis of the data collected.

2.2. Estimating and Testing MTMM Models Based on SB-MTMM Experiments

The main difference with the standard approach is that in the SB-MTMM experiment various samples of the same population, not just one sample, are analyzed simultaneously. Since the samples are drawn from the same population, we assume a common model—the one specified in equations (1) and (2), including the restrictions on the parameters shown in (4a), (4b), and (4c)—even though not all the questions have been asked in every group of respondents. The latter feature of this design is the advantage of this approach: It reduces the response burden for respondents, since the respondents in each sample answer just some of the questions (with the questions being answered differing across groups).

Since the assignment of individuals to groups is made at random, and there is a large sample in each group, the simultaneous analysis of the various groups will be done by using multiple-group SEM (Jöreskog 1971), an approach that is available in most SEM software packages. We refer to this approach as MG-SEM.⁴ As indicated in the previous section, a common model is fitted across the samples, with equality constraints of all parameters across groups. Under current software and theory for multiple-group analysis, estimation can be done by normal theory maximum likelihood (ML) or by any other standard estimation procedure in SEM. In the case of nonnormal data, robust standard errors and test statistics are typically available in the standard software.

Satorra (1993) discusses asymptotic robustness of normal theory methods for multiple-group analysis. He shows that the ML standard

⁴As each group will be confronted with partially different measurements of the same traits, certain software for multiple-group analysis will require some tricks to be applied. This is the case with LISREL, where the standard approach expects the same set of observable variables in each group. Simple tricks to handle such a situation of the set of observable variables differing across groups were already described in the early work of Jöreskog (1971) and in the manual of the early versions of the LISREL program; such procedures are also described in Allison (1987). Multiple-group analysis with the software EQS, for example, does not require the same variables in the different groups. So in EQS we do not need these procedures.

errors of some parameters (loadings and effects parameters), as well as the chi-square goodness-of-fit test statistic, are asymptotically robust to deviations from normality as long as the nonnormal random constituents of the model (error terms, trait, occasion, and method factors) fulfill the following two conditions: (1) unconstrained variances and covariances and (2) mutual independence, not merely zero correlation. In our model setup, however, such conditions do not hold since we impose equality across groups of all model parameters, including the variances and covariances of the trait and method factors (the possible nonnormal constituents of the model); thus in cases of nonnormality standard ML inferences may be wrong. For nonnormal multiple-group data, though, formulas to robustify standard errors and test statistics to deviations from normality are available in standard software. For a review of multiple-group analysis of SEM models that applies to all the designs considered in the present paper and under different distributional conditions see Satorra (2001).

The incomplete data setup we are facing could also be seen as a missing data problem. In the case of a limited number of missing patterns, such as those found in our setup, normal theory ML estimation for missing data was investigated by Muthén, Kaplan, and Hollis (1987), who showed that the same fitting function could be used in this case as in normal theory ML multiple-group approach with means included in the analysis. That is, under our design, the missing-data approach under the normality assumption gives identical results to the ML multiple-group option of analysis just described. In fact, ML for missing data has recently become available in some SEM software programs, so we could just use the option of SEM with missing data (normal theory) to achieve the same results as the normal theory multiple-group option. Note, however, that the missing-data approach typically assumes normality, and it is not yet known how good these procedures are in case of nonnormality. Furthermore, the missing-data approach requires the inclusion of means in the analysis, something that with MG-SEM can be avoided. The adjustment of the analysis when some of the variables are categorical is also straightforward when MG-SEM is used, following the classic approach of Muthén (1984) for categorical ordinal data.

Since the MG-SEM approach provides all the statistics needed, even the ones protected against nonnormality and the approach for categorical data, we advocate the MG-SEM approach as the standard

method of analysis for the SB-MTMM model. In it, the covariance matrices are used as matrices to be analyzed while the data quality criteria, reliability, validity coefficients, and method effects are obtained by complete standardization of the solution obtained.

Although the statistical literature suggests that the data quality indicators discussed above can be estimated through SB-MTMM designs, it cannot be excluded that the two-group design with incomplete data, in particular, may lead to problems due to empirical under-identification. Before discussing these issues, we will first illustrate the use of the two- and three-groups MTMM designs on the basis of the data from the same study discussed at the beginning of this paper.

3. EMPIRICAL EXAMPLES

In Section 1, an empirical example of a standard MTMM experiment was discussed. To illustrate the difference between this design and the SB-MTMM designs, we randomly split the total sample of that study ($n = 428$) in two ($n = 210$) and three groups ($n = 140$). Following this, we took from the full set of observed variables for each group only those variables that would have been collected had the two- or three-group MTMM design been used. In this way, we obtained for each group incomplete covariance matrices. Next, we estimated the model discussed above using the multiple-group approach. We discuss the results in sequence, starting with the three-group design, in which the complete covariance matrix is available from the different groups. We then discuss the results for the two-group design, in which the covariance information is also incomplete.

3.1. *Results for the Three-Group Design*

The random sampling of the different groups and the selection of the variables according to the three-group design led to the results summarized in Table 5.

First, this table indicates that in each sample incomplete data are obtained for the MTMM matrix. The correlations for the unobserved variables are represented by zeros and the variances by ones. This presentation is necessary for the multiple-group analysis with incomplete data in LISREL, but it does not have to be used in general.

TABLE 5
Data for the Three-Group SB-MTMM Analysis on the Basis of Three Random Samples from British Pilot Study of the ESS: Correlations, Means, and Standard Deviations.

First Subsample									
	q1m1	q2m1	q3m1	q1m2	q2m2	q3m2	q1m3	q2m3	q3m3
q1m1	1.00								
q2m1	.469	1.00							
q3m1	.393	.605	1.00						
q1m2	-.669	-.454	-.489	1.00					
q2m2	-.512	-.669	-.564	.707	1.00				
q3m2	-.495	-.508	-.742	.693	.729	1.00			
q1m3	.0	.0	.0	.0	.0	.0	1.00		
q2m3	.0	.0	.0	.0	.0	.0	.0	1.00	
q3m3	.0	.0	.0	.0	.0	.0	.0	.0	1.00
Means	2.41	2.65	2.50	5.18	4.32	4.99	.0	.0	.0
St.dev.	.78	.77	.90	2.39	2.39	2.53	1.0	1.0	1.0

Second Subsample									
	q1m1	q2m1	q3m1	q1m2	q2m2	q3m2	q1m3	q2m3	q3m3
q1m1	1.00								
q2m1	.0	1.00							
q3m1	.0	.0	1.00						
q1m2	.0	.0	.0	1.00					
q2m2	.0	.0	.0	.598	1.00				
q3m2	.0	.0	.0	.601	.694	1.00			
q1m3	.0	.0	.0	.588	.398	.517	1.00		
q2m3	.0	.0	.0	.395	.690	.504	.547	1.00	
q3m3	.0	.0	.0	.397	.462	.571	.545	.564	1.00
Means	.0	.0	.0	5.22	4.30	4.98	1.91	1.69	2.00
St.dev.	1.0	1.0	1.0	2.27	2.51	2.47	.69	.65	.71

Third Subsample									
	q1m1	q2m1	q3m1	q1m2	q2m2	q3m2	q1m3	q2m3	q3m3
q1m1	1.00								
q2m1	.469	1.00							
q3m1	.250	.415	1.00						
q1m2	.0	.0	.0	1.00					
q2m2	.0	.0	.0	.0	1.00				

q3m2	.0	.0	.0	.0	.0	1.00			
q1m3	-.524	-.322	-.212	.0	.0	.0	1.00		
q2m3	-.313	-.523	-.273	.0	.0	.0	.509	1.00	
q3m3	-.244	-.313	-.517	.0	.0	.0	.442	.461	1.00
Means	2.39	2.69	2.41	.0	.0	.0	2.09	1.77	2.02
St. dev.	.70	.71	.78	1.0	1.0	1.0	.71	.68	.73

Note: Zero means and correlations and unit standard deviations represent information that is missing in each group; qimj means question in trait i measured with method j.

It will be clear that these correlation matrices are rather incomplete because, in each of the samples, one set of variables is missing.

Second, we can see that we summarized the response distributions in means and standard deviations and these can be compared across groups, as is done in the standard split-ballot experiments. However, in this case we want more. We also want estimates of the reliability, validity, and method effects. In estimating these coefficients from the data for the three randomly selected groups simultaneously, we assumed that the model is the same for all groups except for the specification of the selection of the variables of the three groups. For the technical details of this analysis, we refer to the input of the LISREL program given in Appendix B. Table 6 provides the results of this estimation as provided by LISREL using the ML estimator.⁵ The table also gives the estimates from the complete data set for comparison.

Given that sampling fluctuations are likely to lead to differences between the different groups, the similarity between the results for the two designs indicates that the three-group SB-MTMM design gives estimates of the parameters of the MTMM model that are very close to the estimates of the classic design, even though the matrices are rather incomplete because people answer fewer questions on the same topic.

LISREL did not face identification problems, even though the covariance matrices in the different subgroups are incomplete. Identification issues are discussed further in Section 4.1. Let us now investigate the same example in the same way, assuming that a two-group design has been used.

⁵In this case LISREL reports a chi-square of 54.7 with d.f. = 111. However, the number of degrees of freedom is incorrect because in each matrix 24 correlations and variances were missing; therefore, the d.f. should be reduced by $3 \times 24 = 72$, making the correct d.f. = 39. Note that all fit indices have to be corrected accordingly for the correct d.f.

TABLE 6
The Estimates of the Parameters for the Full Sample Using Three Methods and
for the Three-Group Design with Incomplete Data in Each Group^a

Reliability	Full Sample			Three-Group SB-MTMM Design		
	M1	M2	M3	M1	M2	M3
Q1	.79	.91	.82	.80	.91	.84
Q2	.85	.94	.87	.87	.97	.86
Q3	.81	.93	.84	.78	.95	.77
Validity						
Q1	.93	.91	.85	.94	.91	.86
Q2	.94	.92	.87	.94	.93	.85
Q3	.95	.93	.88	.96	.93	.84
Method var	.05	.73	.09	.04 ^b	.73	.09

^a Tables 6 and 8 provide the same information for the full sample as Table 2 but in a more compact way.

^b This coefficient is not significantly different from zero; all others are significantly different from zero.

3.2. Two-Group SB-MTMM Design

Using the two-group design, the same model is assumed to hold for both groups and the analysis is carried out in exactly the same way. The data for this design are shown in Table 7.

The procedure for filling in the empty cells was the same in Table 7 as in Table 5. An important difference between the two designs is that in this case no correlations at all are available between the traits measured with the first and the second method. Therefore, the parameters have to be estimated on the basis of an incomplete covariance matrix.

The analysis with this data did converge, but an improper solution was obtained with negative variances for the variances of the first two method factors. This problem also arises in the classic MTMM approach when a method factor has a variance very close to zero. Table 6 shows that the method variance for the first factor was not significantly different from zero and is rather small, even though the estimate was based on two groups of 140 or 280 cases. In the two-group design, this variance has to be estimated on the basis of 210 cases, and in this case it seems that it does not give a proper solution. A common

TABLE 7
The Data for the Two-Group SB-MTMM Analysis on the Basis of Two Random Samples from British Pilot Study of the ESS: Correlations, Means and Standard Deviations.

First Subsample									
	q1m1	q2m1	q3m1	q1m2	q2m2	q3m2	q1m3	q2m3	q3m3
q1m1	1.00								
q2m1	.457	1.00							
q3m1	.347	.478	1.00						
q1m2	.0	.0	.0	1.00					
q2m2	.0	.0	.0	.0	1.00				
q3m2	.0	.0	.0	.0	.0	1.00			
q1m3	-.564	-.365	-.344	.0	.0	.0	1.00		
q2m3	-.366	-.597	-.359	.0	.0	.0	.546	1.00	
q3m3	-.350	-.386	-.530	.0	.0	.0	.512	.498	1.00
Means	2.42	2.75	2.43	.0	.0	.0	2.01	1.70	1.99
St.dev.	.74	.76	.83	1.0	1.0	1.0	.71	.67	.73

Second Subsample									
	q1m1	q2m1	q3m1	q1m2	q2m2	q3m2	q1m3	q2m3	q3m3
q1m1	1.00								
q2m1	.0	1.00							
q3m1	.0	.0	1.00						
q1m2	.0	.0	.0	1.00					
q2m2	.0	.0	.0	.686	1.00				
q3m2	.0	.0	.0	.669	.742	1.00			
q1m3	.0	.0	.0	.585	.449	.441	1.00		
q2m3	.0	.0	.0	.464	.684	.546	.568	1.00	
q3m3	.0	.0	.0	.397	.516	.674	.516	.607	1.00
Means	.0	.0	.0	5.26	4.49	5.10	2.01	1.80	2.02
St.dev.	1.0	1.0	1.0	2.38	2.40	2.51	.74	.73	.81

Note: Zero means and correlations and unit standard deviations represent information that is missing in each group; qimj means question in trail i measured with method j.

solution in such cases is to fix one parameter at a value close to zero. If we fix this variance at .01, we get the result presented in Table 8.⁶

⁶In this case LISREL reports a chi-square value of 12.7 with d.f. = 67, but here too the d.f. has to be corrected in the way explained above (note 5) making the correct d.f. = 19.

TABLE 8
The Estimates of the Parameters for the Full Sample Using Three Methods and
for the Two-Group Design with Incomplete Data in Each Group

Reliability	Full Sample			Two-Group SB-MTMM Design		
	M1	M2	M3	M1	M2	M3
Q1	.79	.91	.82	.80	.93	.83
Q2	.85	.94	.87	.87	.96	.86
Q3	.81	.93	.84	.83	.98	.82
Validity						
Q1	.93	.91	.85	.99	.90	.85
Q2	.94	.92	.87	.99	.91	.86
Q3	.95	.93	.88	.99	.92	.87
Method variances	.05	.73	.09	.01 ^a	.86	.10

^aThis coefficient was fixed on the value .01 in order to avoid an improper solution.

Table 8 shows that, with the restriction discussed above, the program provides estimates that are not too far from the estimates used in the classic MTMM design. The largest differences in the validity coefficients for the first method are a direct consequence of the restriction introduced. On the whole, the conclusion drawn from the estimates obtained by the two-group design does not differ from the conclusion drawn from the estimates of the one-group design: the second method is more reliable. Given the restriction introduced on one method variance, we would be reluctant to draw a definite conclusion about the validity coefficients and therefore about the method effects.

Clearly, the fact that we had to introduce this restriction raises the question of whether the two-group design is identified and whether it is stable enough to be useful in practice. On the one hand, it would seem to be the most natural approach to reducing the response burden. On the other hand, when this approach is not stable enough to provide the same estimates as the classic or the three-group SB-MTMM design, then one of the other designs has to be preferred.

Regarding identification, we can say that the model is indeed identified under normal circumstances and the estimation procedure specified will provide consistent estimates of the population parameters. This can be verified by assessing the full rank of the Jacobian matrix associated with the model specified. This issue will be discussed in Section 4.1.

Before proceeding to the next section, it should be noted that the example above did not give a correct impression of the quality of the different designs. The reason is that the quantity of data on the basis of which the parameters were estimated differed for the parameters in the different designs. The parameters of the classic design were based on approximately 420 cases. The parameter estimates in the three-group design are based on 140 or 280 respondents. In the two-group design the parameter estimates are based on 210 or 420 cases. These differences in sample sizes could be a reason for difference in performance. We therefore also discuss the topic of efficiency of the different designs in the next section.

4. THE EMPIRICAL IDENTIFIABILITY AND EFFICIENCY OF DIFFERENT SB-MTMM DESIGNS

To assess the empirical performance of these different designs, two issues have to be investigated. The first is under what conditions the procedures break down, even though the correct model has been specified. The second issue concerns the efficiency of the designs in estimating the parameters of the MTMM model. We begin with the first issue.

4.1. *The Empirical Identifiability of the SB-MTMM Model*

There are three aspects of these models that require special attention when the model has been specified correctly:

- Minimal variance of one of the method factors
- Lack of correlation between the latent traits
- Equal correlations between the latent traits

The problem of minimal method variance is a problem of overfitting. In this case a parameter is estimated that is not needed to fit the model to the data. If the model had been estimated with this coefficient set at zero, the fit would be equally good. This problem is not just a problem of SB-MTMM designs; it also occurs in the classic MTMM design. The solution, as mentioned above, is to specify the parameter that is not needed for the model at zero or close to zero.

It is more problematic to detect where the problem in the model is. Our experience with analyses of MTMM data is that negative variances for the method variances are obtained in unrestricted estimation procedures if the variances are very close to zero. In such cases, restricting the variances to a value very close to zero solves the problem. In the case where estimation procedures include constraints on the parameter values in order to avoid improper solutions, the value zero will automatically be obtained for the problematic method variance.

The second condition, lack of correlations between the traits, can create a problem because it is known that the loadings of a factor model are identified if each trait has three indicators, or if each trait has two indicators and the traits are correlated with each other. If each trait has only two indicators and the correlation between the traits is zero, the situation is the same as for a nonidentified model with one trait and two indicators. Applying this rule to the MTMM models, we can see that in the classic MTMM model each trait has three indicators and is therefore under normal circumstances identified even if the correlations between the traits are zero. In the different groups of the SB-MTMM designs, each trait has only two indicators. Therefore, if the correlation between two traits is zero, the model in the different subgroups will not be identified. If all three correlations, or two of the three, go to zero, the standard errors of the parameters become very large. This is an indication that a problem of identification exists.

Fortunately, there is a simple solution to this problem if we have some freedom of choice in the selection of the traits for the experiments. We can then select as traits for the experiment those traits that have sufficient correlation to avoid problems. If we are aware of this problem, we can prevent it in the design of the experiment.

The third condition is that the basic model of the two-group SB-MTMM design is not identified if the correlations between the traits are exactly identical. Fortunately, this is a very unlikely situation. However, if we are confronted with a situation where the standard errors are rather large while the correlations between the traits are not close to zero, equality of the correlations might be the explanation.

The discussion so far suggests that the SB-MTMM design with two groups can be used if we select traits that correlate with each other but do not have equal correlations. Under these rather elementary conditions, even the two-group SB-MTMM designs will be identified

and the multiple-group ML estimator will provide consistent estimates. For the three-group design, these requirements are not necessary.

4.2. *The Efficiency of the Various Designs*

The second issue to be discussed is the efficiency of the various designs. This is a relevant issue because the reduction of the response burden might be gained at the expense of efficiency. Efficiency is here studied on the basis of the standard errors of the estimates of reliability and validity. Given that reliability and validity are estimated as two standardized parameters, and the standard errors of standardized coefficients are not available in most SEM programs, the procedure for computing the standard errors is provided in Appendix C.

Efficiency will be evaluated by determining the total sample size over the two or three groups needed in each design to obtain the same standard error for the relevant parameters, as in the classic one-group design. As a starting point for the evaluation of efficiency, we chose an analysis of one-group design with a sample size of 300 cases. This sample size is chosen because it is normally sufficiently accurate. The data for analysis were generated with a model in which all the method variances are equal while the validity coefficients squared plus the method variances are equal to 1 and the error variances are also equal to each other for all nine variables. This was done to simplify the results. In such a model, only two parameters have to be chosen:⁷ method variance and error variance. The upper part of Figure 3 gives the sample sizes needed (for each of the groups) in the two- and three-group design to obtain the same precision in estimation of validity as in the one-group design ($n = 300$) for different variances of the method effect for a fixed value of error variance (.30).⁸ The lower part of Figure 3 gives the sample sizes needed (for each of the groups) in the two- and three-group design to obtain the same precision in estimation of reliability as in the one-group design ($n = 300$) for

⁷The correlations between the traits are also parameters of the model, but these parameters have not been varied as the other two have. The values of these parameters were .6 for traits 1 and 2; .3 for traits 1 and 3; and .1 for traits 2 and 3.

⁸Because the estimates of the parameters of the two-group design are based on different numbers of parameters, we use here the worst case—i.e. the result for parameters based on only one group in the two-group design.

FIGURE 3. The sample size needed in the two- and three-group design to obtain the same accuracy in estimation of the reliability and validity as in the one-group design ($n = 300$) for different variances of the method effect and the random errors.

different variances of the random errors for a fixed value of method variance (.16). The fixed values represent reasonable values of the parameters in practice.

Figure 3 shows that for very small method variances the total sample for the two-group design has to be very large. A much smaller total sample is needed for the three-group design. However, we should realize that the standard error for very small method variances is also very small unless the variance is equal to zero, as discussed above.

This figure also shows the same kind of results for the effect of the error variance on the sample size required in two- and three-groups designs if the same precision is to be obtained in the estimation of reliability as in one-group design.

The inefficiency of the two designs for very small error variances, compared with the one-group design, is also evident.

Fortunately—or unfortunately—these very small error variances do not occur in survey research. This figure shows clearly the price we have to pay for the reduction of the response burden in the two- and three-group design.

5. CONCLUSION AND DISCUSSION

This paper has shown that the SB-MTMM experiment provides the same information as the more common split-ballot design on the distribution of responses for different forms of the same question. But the SB-MTMM design can also provide information about the reliability and validity of measurements if we are willing to ask three more questions in each group. This is an important advantage over the standard split-ballot design.

Compared with the classic one-group MTMM design, the SB-MTMM design reduces response burden by reducing the number of items to be asked in a questionnaire, without loss of information on reliability and validity measurements. Questions concerning the same trait need to be answered only twice, not three times as is required in the classic MTMM approach. Thus its major advantage is that it reduces the response burden. It is, however, also clear that a price is paid for this design improvement. The sample size required in SB-MTMM designs is much larger than in one-group designs, as was shown in Section 4.

It should be noted that the effects of repeating questions dealing with the same concept cannot be eliminated completely. Repetition is necessary for estimating the reliability and validity. However, occasion-specific effect or order effects can be estimated using designs with repeated observations of the same traits with exactly the same questions. Fortunately, this does not have to be done for all forms and traits. Three- or four-group designs with exactly repeated observations for one method are sufficient to estimate these effects. Meta-analyses of MTMM experiments also provide estimates of the effect of repeated observations and allow correction for this effect, as has been shown by Saris and Gallhofer (forthcoming).

For estimation, we suggest analyzing the data of these multiple-group designs by using the options of multiple-group SEM available in

standard software. It is important to note in this context that we can obtain corrections to standard errors and test statistics to cope with nonnormality in a standard fashion.

Regarding efficiency, we have shown that the three-group design is far more efficient than two-group design, especially for small method variances and error variances. The total sample sizes can be reduced by the use of three groups instead of two if the errors are quite small. However, the three-group design also has disadvantages—for example, it does not give data for the same variables for all people in the sample. At least one group will have incomplete data. In the two-group design, this is not the case, because all respondents are confronted for each trait with one of the three forms.

Another disadvantage is that the three-group design requires more forms of the questionnaire. This may create problems in paper and pencil research if the designs become more complex. The decision about which design should be used in practice will depend on the design of the study. Let us illustrate this point.

Comparing the two-group and the one-group design, we can observe the following. For an averaged survey item with a method variance around .16 and a averaged error variance around .3, the standard error for reliability and validity is close to .03 in a one-group design with a sample size of 300. To get the same accuracy with a two-group design, we need at least 700 cases in each group. In a study with 1500 cases, we could do 5 one-group MTMM studies but only 1 two-group design study. If the one-group design is used, this means that each group gets 3 questions, which have to be answered three times. In two-group design, no group has to answer the same question three times, and so each group has three fewer questions to answer than in the one-group design. Therefore, we could also use each group in the two-group design for two experiments. Each group would then have to answer the same number of questions as in the one-group design, but none of the questions would be asked more than twice. Consequently, the comparison is between 5 one-group MTMM experiments with three questions that have to be asked three times, and 2 two-group MTMM experiments with an equal number of questions, but none of the questions has to be asked more than twice. Accuracy would be approximately the same, as would be the complexity of the field work, but the problems of repeating questions three times would be avoided in the two-group design.

Depending on the number of questions we would like to evaluate in a MTMM study, and the size of the sample in which the MTMM experiments are to be placed, we have to study what the most efficient way is to use these designs within a specific project. Further discussion of this issue would be beyond the scope of this paper. We wish only to show here that there is an alternative to the classic MTMM design: the SB-MTMM design.

APPENDIX A: THE RELATIONSHIP BETWEEN THE TS MODEL AND THE CLASSIC MTMM MODEL

The structure of the classic MTMM model follows directly from the basic characteristics of the TS model that has already been specified in equations (1) and (2) above. From this model we can derive the most commonly used MTMM model by substitution of equation (2) into equation (1). The result is the model

$$Y_{ij} = r_{ij}v_{ij}F_i + r_{ij}m_{ij}M_j + e_{ij} \quad (\text{A} - 1)$$

or

$$Y_{ij} = q_{ij}F_i + s_{ij}M_j + e_{ij}, \quad (\text{A} - 2)$$

where $q_{ij} = r_{ij}v_{ij}$ and $s_{ij} = r_{ij}m_{ij}$. One advantage of this formulation is that q_{ij} gives the strength of the relationship between the variable of interest and the observed variable and is as such an important indicator of quality of an instrument, while s_{ij} gives the systematic effect of method j on response Y_{ij} . Another advantage is that it simplifies equation (3) to

$$r(R_{1j}, R_{2j}) = q_{1j}r(F_1, F_2)q_{2j} + s_{1j}s_{2j} \quad (\text{A} - 3)$$

Although this model looks very attractive, there are some problems associated with it. One is that the estimates of data quality for any model are obtained only after an MTMM experiment has been conducted and the data analyzed. In order to apply this approach in practice, for each question in the survey we should ask two more questions to estimate quality. This is of course prohibitively expensive and therefore not done.

An alternative would be to study the effects of different questionnaire design choices on quality criteria and use these relationships to predict data quality before and after data are collected. If enough MTMM experiments are carried out and a meta-analysis to determine the effects of choices on quality criteria is undertaken, then no extra questions are needed in the substantive surveys. This approach is indeed what has been suggested by Andrews (1984) and is also applied in several other studies (Költringer 1995; Scherpenzeel and Saris 1997; Saris and Gallhofer, forthcoming).

However, in such an analysis of quality criteria, it is preferable to use parameter estimates that represent only one criterion and not mixtures of different criteria that could confuse the explanation. It is for this reason that Saris and Andrews have suggested an alternative parameterization of the classic model. This True Score model is already seen in equations (1) and (2). In this model, the reliability and validity coefficients are separated and can be estimated independently of each other. Both can also vary between 0 and 1, which is not true if we use the reliability and the coefficient q_{ij} (as Andrews [1984] did) starting with the classic model. Saris and Andrews (1991) have suggested that for meta-analysis the True Score MTMM model has major advantages. Since we think that meta-analysis across MTMM experiments is the most important application of the MTMM design, we use the True Score model in this paper.

APPENDIX B: THE LISREL INPUT FOR THE THREE-GROUP SB-MTMM EXAMPLE

```
Analysis of british satisfaction data with 3groups SB-MTMM model
group 1
Data ng=3 ni=9 no=140 ma=cm
km
1.000
0.469 1.000
0.250 0.415 1.000
0.000 0.000 0.0000 1.000
0.0000 0.0000 0.0000 0.0000 1.000
0.0000 0.0000 0.0000 0.0000 0.0000 1.000
-.524 -.322 -.212 0.0000 0.0000 0.0000 1.000
-.313 -.523 -.273 0.0000 0.0000 0.0000 0.509 1.000
-.244 -.313 -.517 0.0000 0.0000 0.0000 0.442 0.461 1.000
```

```

me
2.39 2.69 2.41 0.0 0.0 0.0 0.0 2.09 1.77 2.02
sd
.70 .71 .78 1.0 1.0 1.0 .71 .68 .73
model ny=9 ne=9 nk=6 ly=fu, fi te=di, fr ps=di, fi be=fu, fi
ga=fu, fi ph=sy, fi
value -1 ly 1 1 ly 2 2 ly 3 3
value 0 ly 4 4 ly 5 5 ly 6 6
pa te
15 16 17 0 0 0 18 19 20
value 1 te 4 4 te 5 5 te 6 6
value 1 ly 7 7 ly 8 8 ly 9 9
free ga 1 1 ga 4 1 ga 7 1 ga 2 2 ga 5 2 ga 8 2 ga 3 3 ga 6 3 ga 9 3
value -1 ga 1 4 ga 2 4 ga 3 4
value 1 ga 4 5 ga 5 5 ga 6 5 ga 7 6 ga 8 6 ga 9 6
free ph 2 1 ph 3 1 ph 3 2 ph 4 4 ph 5 5 ph 6 6
start .01 ph 4 4
value 1 ph 1 1 ph 2 2 ph 3 3
start .5 all
value .13 ph 5 5
value .18 ph 6 6
start .75 ga 1 1 ga 2 2 ga 3 3 ga 7 1 ga 8 2 ga 9 3
start .85 ga 4 1 ga 5 2 ga 6 3
out iter=200 adm=off sc

```

Analysis of british satisfaction group 2

Datani=9 no=150 ma=cm

Km

```

1.000
0.0    1.000
0.0    0.0    1.000
0.0    0.0    0.0    1.000
0.0    0.0    0.0    0.598    1.000
0.0    0.0    0.0    0.601    0.694    1.000
0.0    0.0    0.0    0.588    0.398    0.517    1.000
0.0    0.0    0.0    0.395    0.690    0.504    0.547    1.000
0.0    0.0    0.0    0.397    0.462    0.571    0.545    0.564    1.000

```

```

me
.0 .0 .0 5.22 4.30 4.98 1.91 1.69 2.00
sd
1.0 1.0 1.0 2.27 2.51 2.47 .69 .65 .71
model ny=9 ne=9 nk=6 ly=fu, fi te=di, fr ps=in be=in ga=in
ph=in
value 0 ly 1 1 ly 2 2 ly 3 3
pa te
0 0 0 21 22 23 18 19 20

```

```

value 1 te 1 1 te 2 2 te 3 3
value 1 ly 4 4 ly 5 5 ly 6 6 ly 7 7 ly 8 8 ly 9 9
out iter=200 adm=off sc

Analysis of british satisfaction group 3
Data ni=9 no=150 ma=cm
Km
*
1.000
0.469 1.000
0.393 0.605 1.000
-.669 -.454 -.489 1.000
-.512 -.669 -.564 .707 1.000
-.495 -.508 -.742 .693 .729 1.000
0.0 0.0 0.0 .000 .000 0.000 1.000
0.0 0.0 0.0 .000 .000 0.000 0.000 1.000
0.0 0.0 0.0 0.0000 0.0000 0.000 0.000 0.000 1.000
me
2.41 2.65 2.50 5.18 4.32 4.99 .0 .0 .0
sd
.78 .77 .90 2.39 2.39 2.53 1.00 1.00 1.00
model ny=9 ne=9 nk=6 ly=fu, fi te=di, fr ps=in be=in ga=in
ph=in
value 0 ly 7 7 ly 8 8 ly 9 9
pa te
15 16 17 21 22 23 0 0 0
value 1 te 7 7 te 8 8 te 9 9
value 1 ly 4 4 ly 5 5 ly 6 6
value -1 ly 1 1 ly 2 2 ly 3 3
out iter=200 adm=off sc

```

APPENDIX C: STANDARD ERRORS OF RELIABILITY AND VALIDITY ESTIMATES

This appendix provides the expressions for the standard errors of the estimates of reliability and validity in the SB-MTMM model. The standard errors are computed as the square root of the asymptotic variances of the reliability and validity estimates derived using the classical delta method. The reliability and validity coefficients, r^2 and v^2 , can be expressed as functions of basic parameters of the SB-MTMM model. The basic model used to estimate the parameters is represented by the following two equations:

$$Y_{ij} = T_{ij} + e_{ij} \quad \text{for } i = 1-3 \text{ and } j = 1-3 \quad (\text{c} - 1)$$

$$T_{ij} = \lambda_{ij}F_i + M_j \quad \text{for } i = 1-3 \text{ and } j = 1-3 \quad (\text{c} - 2)$$

where $\text{var}(F_i) = 1$; $\text{var}(M_j) = \phi_j$; and $\text{var}(e_{ij}) = \theta_{ij}$.

We can then define reliability and validity as functions of the parameters of this model:

$$r_{ij}^2 = \text{var}(T_{ij})/\text{var}(Y_{ij}) = \text{gr}(\lambda_{ij}, \phi_j, \theta_{ij}) = (\lambda_{ij}^2 + \phi_j)/(\lambda_{ij}^2 + \phi_j + \theta_{ij}) \quad (\text{c} - 3)$$

$$v_{ij}^2 = \text{var}(F_i)/\text{var}(T_{ij}) = \text{gv}(\lambda_{ij}, \phi_j) = \lambda_{ij}^2/(\lambda_{ij}^2 + \phi_j), \quad (\text{c} - 4)$$

where $\text{gr}(\cdot)$ and $\text{gv}(\cdot)$ are continuously differentiable functions.

Since standard computer software such as LISREL or EQS provides estimates of the vector of parameters $(\lambda_{ij}, \phi_j, \theta_{ij})$ and a corresponding asymptotic variance-covariance matrix of the estimates, straightforward application of the delta method produces the following expressions for the variances of the estimates of v^2 and r^2 :

$$\text{var}(\text{estimate of } v^2) = \text{dgv } \mathbf{V}_1 (\text{dgv})' \quad (\text{c} - 5)$$

$$\text{var}(\text{estimate of } r^2) = \text{dgr } \mathbf{V}_2 (\text{dgr})' \quad (\text{c} - 6)$$

where

$$\begin{aligned} \text{dgv} &= [\partial \text{gv} / \partial \lambda_{ij}, \partial \text{gv} / \partial \phi_j] = [2\lambda_{ij}\phi_j/(\lambda_{ij}^2 + \phi_j)^2, -\lambda_{ij}^2/(\lambda_{ij}^2 + \phi_j)^2] \\ \text{dgr} &= [\partial \text{gr} / \partial \lambda_{ij}, \partial \text{gr} / \partial \phi_j, \partial \text{gr} / \partial \theta_{ij}] \\ &= [2\lambda_{ij}\theta_{ij}/(\lambda_{ij}^2 + \phi_j + \theta_{ij})^2, \theta_{ij}/(\lambda_{ij}^2 + \phi_j + \theta_{ij})^2, \\ &\quad -(\lambda_{ij}^2 + \phi_j)/(\lambda_{ij}^2 + \phi_j + \theta_{ij})^2] \end{aligned}$$

and $\mathbf{V}_1$ is the variance-covariance matrix of the estimates of the vector of parameters (λ_{ij}, ϕ_j) and $\mathbf{V}_2$ is the variance-covariance matrix of the estimates of the vector of parameters $(\lambda_{ij}, \phi_j, \theta_{ij})$.

The square roots of the above expression of $\text{var}(\text{estimate of } v^2)$ and $\text{var}(\text{estimate of } r^2)$ are the desired (asymptotic) standard errors that are used to construct the graphs of Section 4.2. We should emphasize that these are standard errors whose validity is sustained by the large sample size assumption, and for the condition of the variance $\lambda_{ij}^2 + \phi_j + \theta_{ij}$ not being too small.

REFERENCES

- Allison, P. D. 1987. "Estimation of Linear Models with Incomplete Data." Pp. 71–103 in *Sociological Methodology*, Vol 17, edited by C. C. Clogg. Washington, DC: American Sociological Association.
- Althausen, R. P., T. A. Heberlein, and R. A. Scott. 1971. "A Causal Assessment of Validity: The Augmented Multitrait-Multimethod Matrix." Pp. 151–69 in *Causal Models in the Social Sciences*, edited by H. M. Blalock, Jr. Chicago: Aldine.
- Alwin, D. 1974. "An Analytic Comparison of Four Approaches to the Interpretation of Relationships in the Multitrait-Multimethod Matrix." Pp. 79–105 in *Sociological Methodology*, edited by H. L. Costner. San Francisco: Jossey-Bass.
- Andrews, F. M. 1984. "Construct Validity and Error Components of Survey Measures. A Structural Modeling Approach." *Public Opinion Quarterly* 48:409–42.
- Arminger, G., and M.E. Sobel. 1990. "Pseudo-Maximum Likelihood Estimation of Mean and Covariance Structures with Missing Data." *Journal of the American Statistical Association* 85:195–203.
- Bagozzi, R. P., and Y. Yi. 1991. "Multitrait-Multimethod Matrices in Consumer Research." *Journal of Consumer Research* 17:426–39.
- Billiet, J., G. Loosveldt, and L. Waterplas. 1986. "Het Survey-Interview Onderzocht: Effecten van het Ontwerp en Gebruik van Vragenlijsten op de Kwaliteit van de Antwoorden" (Research on surveys: effects of the design and use of questionnaires on the quality of the responses). Leuven, Belgium: Sociologisch Onderzoeksinstituut KU Leuven.
- Brannick, M. T., and P. E. Spector. 1990. "Estimation Problems in the Block-Diagonal Model of the Multitrait-Multimethod Matrix." *Applied Psychological Measurement* 14:325–39.
- Browne M. W. 1984. "The Decomposition of Multitrait-Multimethod Matrices." *British Journal of Mathematical and Statistical Psychology* 37:1–21.
- Bunting, B., G. Adamson, and P. K. Mulhall. 2002. "A Monte Carlo Examination of an MTMM Model with Planned Incomplete Data Structures." *Structural Equation Modeling* 9:369–89.
- Campbell, D. T., and D. W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait Multimethod Matrices." *Psychological Bulletin* 56: 81–105.
- Campbell, D. T., and E. J. O'Connell. 1967. "Method Factors in Multitrait-Multimethod Matrices: Multiplicative Rather Than Additive?" *Multivariate Behavioral Research* 2:409–26.
- Coenders, G., and W. E. Saris. 1995. "Categorization and Quality: The Choice Between Pearson and Polychoric Correlations." Pp. 125–44 in *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, edited by W. E. Saris and A. Münnich. Budapest: Eötvös University Press.
- Coenders, G., A. Satorra, and W. E. Saris. 1997. "Alternative approaches to Structural Modeling of Ordinal Data: A Monte Carlo Study." *Structural Equation Modeling* 4:261–82.

- . 1998. "Relationship Between a Restricted Correlated Uniqueness Model and a Direct Product Model for Multitrait-Multimethod Data." Pp. 151–72 in *Advances in Methodology, Data Analysis and Statistics, Metodoloki Zvezki*, Vol 14. Ljubljana, Slovenia.
- . 2000. "Testing Nested Additive, Multiplicative and General Multitrait-Multimethod Models." *Structural Equation Modeling* 7:219–50.
- Corten I., W. E. Saris, G. Coenders, W. van der Veld, C. Albers, and C. Cornelis. 2002. "The Fit of Different Models for Multitrait-Multimethod Experiments." *Structural Equation Modeling* 9:213–32.
- Cudeck, R. 1988. "Multiplicative Models and MTMM Matrices." *Journal of Educational Statistics* 13, 131–47.
- Eid, M. 2000. "Multitrait-Multimethod Model with Minimal Assumptions." *Psychometrika* 65:241–61.
- European Social Survey. 2002. *European Social Survey Round 1: Report of the First Year*. London: Natcen.
- Grayson, D., and H. W. Marsh. 1994. "Identification with Deficient Rank Loading Matrices in Confirmatory Analysis: Multitrait-Multimethod Models." *Psychometrika* 59:121–34.
- Jöreskog, K. G. 1971. "Simultaneous Factor Analysis in Several Populations." *Psychometrika* 34:409–26.
- . 1990. "New Developments in LISREL: Analysis of Ordinal Variables Using Polychoric Correlations and Weighted Least Squares." *Quality and Quantity* 24:387–404.
- Kenny, D. A. 1976. "An Empirical Application of Confirmatory Factor Analysis to the Multitrait-Multimethod Matrix." *Journal of Experimental Social Psychology* 12:247–52.
- Kenny, D. A., and Kashy, D. A. 1992. "Analysis of the Multitrait-Multimethod Matrix by Confirmatory Factor Analysis." *Psychological Bulletin* 112:165–72.
- Kogovšek, T., A. Ferligoj, G. Coenders, and W. E. Saris. 2002. "Estimating 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 Austria Personal Interviews." Pp. 207–25 in *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, edited by W.E. Saris and A. Münnich. Budapest: Eötvös University Press.
- Marsh, H. W. 1989. "Confirmatory Factor Analysis of Multitrait-Multimethod Data: Many Problems and Few Solutions." *Applied Psychological Measurement* 13:335–61.
- Marsh, H. W., and M. Bailey. 1991. "Confirmatory Factor Analyses of Multitrait-Multimethod Data: Comparison of the Behavior of Alternative Models." *Applied Psychological Measurement* 15:47–70.
- van Meurs, A., and W. E. Saris. 1990. "Memory Effects in MTMM Studies." Pp. 134–46 in *Evaluation of Measurement Instruments by Meta-analysis of Multitrait-Multimethod Studies*, edited by W. E. Saris and A. van Meurs. Amsterdam: North Holland.

- Muthén, B. 1984. "A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators." *Psychometrika* 49:115–32.
- Muthén, B., D. Kaplan, and M. Hollis. 1987. "On Structural Equation Modeling with Data That Are Not Missing Completely at Random." *Psychometrika* 52:431–62.
- Olsson, U. 1979. "Maximum-Likelihood Estimation of the Polychoric Correlation Coefficient." *Psychometrika* 44:443–60.
- Rodgers, W. L., F. M. Andrews, and A. R. Herzog. 1992. "Quality of Survey Measures: A Structural Modeling Approach." *Journal of Official Statistics* 8:251–75.
- Saris, W. E. 1990. "The Choice of a Model for Evaluation of Measurement Instruments." Pp. 118–29 in *Evaluation of Measurement Instruments by Meta-analysis of Multitrait Multimethod Matrices*, edited by W. E. Saris and A. van Meurs. Amsterdam: North Holland.
- . 1998. "A New Approach for Evaluation of Measurement Instruments: The Split-Ballot MTMM Design." Presented at the International Conference on Methodology and Statistics, Preddvor, Slovenia, September 20–22, 1998.
- Saris, W. E., and F. M. Andrews. 1991. "Evaluation of Measurement Instruments Using a Structural Modeling Approach." Pp. 575–99 in *Measurement Errors in Surveys*, edited by P. P. Biemer et al. New York: Wiley.
- Saris, W. E. and I. N. Gallhofer. Forthcoming. "Estimation of the Effects of Measurement Characteristics on the Quality of Survey Questions."
- Saris, W. E., and J. Krosnick. Forthcoming. "Comparing Questions with Agree/Disagree Response Options to Questions with Construct Specific Response Options."
- Saris, W. E., and A. Munnich. 1995. *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*. Budapest: Eötvös University Press.
- Saris, W. E., T. Van Wijk, and A. Scherpenzeel. 1998. "Validity and Reliability of Subjective Social Indicators: The Effect of Different Measures of Association." *Social Indicators Research* 45:173–99.
- Satorra, A. 1992. "Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures." Pp. 249–78 in *Sociological Methodology*, Vol. 22, edited by P. V. Marsden. Cambridge, MA: Basil Blackwell.
- Satorra, A. 1993. "Asymptotic Robust Inferences in Multi-Sample Analysis of Augmented-Moment Structures." Pp. 211–29 in *Multivariate Analysis: Future Directions*, Vol. 2, edited by C. M. Cuadras and C. R. Rao. Amsterdam: Elsevier.
- Satorra, A. 2001. "Goodness of Fit Testing of Structural Equation Models with Multiple Group Data and Nonnormality." Chap. 12 in *Structural Equation Modeling: Present and Future*, edited by R. Cudeck, S. du Toit, and D. Sörbom. Lincolnwood, IL: Scientific Software International. SSI.
- Scherpenzeel, A. C. 1995. "A Question of Quality. Evaluating Survey Questions by Multitrait-Multimethod Studies." Ph.D. dissertation, University of Amsterdam, Leidschendam, the Netherlands.

- Scherpenzeel, A., and W. E. Saris. 1997. "The Validity and Reliability of Survey Questions: A Meta-analysis of MTMM Studies." *Sociological Methods and Research* 25:341–83.
- Schuman, H., and S. Presser, 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Order, and Context*. New York: Academic Press.
- Werts, C. E., and R. L. Linn, 1970. "Path Analysis. Psychological Examples." *Psychological Bulletin* 74:193–212.
- Wothke, W. 1996. "Models for Multitrait-Multimethod Matrix Analysis. Pp. 7–56 in *Advanced Structural Equation Modeling. Issues and Techniques*, edited by G. C. Marcoulides and R. E. Schumacker. Mahwah, NJ: Lawrence Erlbaum.