THE HISTORY AND THE BASIC PUBLICATIONS IN THE DEVELOPMENT OF SQP

2015

To cite this document: Survey Quality Predictor (2015). The history and the basic publications in the development of SQP. Barcelona, Universitat Pompeu Fabra.

Table of contents

In 1971 the LISREL program was introduced to the social sciences, which allowed researchers to study models that distinguish between latent variables and observed variables, and in this way allowed for the estimation of the measurement errors in measurement instruments. One of the models developed in this context was the Multitrait-Multimethod (MTMM) model (Jöreskog, 1971) based on the Multitrait-Multimethod design originally developed by Campbell and Fiske (1959).

Applying LISREL to social science data, it became clear that the measurement errors are considerable. Alwin (2007) estimated the error variance on half of the variance of observed variables based on survey questions.

The consequence of the errors in survey questions is that the estimates of correlations and regression coefficients are biased, unless correction for measurement errors is applied (Saris and Gallhofer, 2014). Without correction for measurement errors, the strength of relationships between variables cannot be compared and comparison of relationships across countries is not possible (see e.g. Oberski, 2011). A logical reaction is to try to improve measurements but errors will nevertheless always remain. Therefore, throughout the following years different approaches to deal with the errors were studied:

PHASE 1: THE FIRST MTMM STUDIES

A new phase started when Frank Andrews applied the MTMM design, model and analysis to the evaluation of the quality of single questions. He reported on his research in 1984. He carried out several experiments and a meta-analysis of different experiments to determine the effects of different characteristics and designs on the survey questions.

A typical example of a MTMM experiment is to have three questions which measure the same latent construct; for example, if the construct is satisfaction with government performance in Britain, we could have the following three questions (traits):

> *Trait 1: On the whole, how satisfied are you with the present state of the economy in Britain?*
>
> *Trait 2: Now think about the national government. How satisfied are you with the way it is doing its job?*
>
> *Trait 3: And on the whole, how satisfied are you with the way democracy works in Britain?*

We could also be interested in testing a 'method', for example how measurement quality varies depending on how the question's response scale is designed. For example, we could test three alternative response scales (see Table 1).

Table 1: Three alternative response scales

Alternative 1	Alternative 2	Alternative 3
1) Very satisfied 2) Fairly satisfied 3) Fairly dissatisfied 4) Extremely dissatisfied	Dissatisfied Satisfied 0 1 2 3 4 5 6 7 8 9 10	1) Not at all satisfied 2) Satisfied 3) Quite satisfied 4) Very satisfied

Presenting each question in three different forms to the same respondents, one obtains a correlation matrix with 45 variances and covariances. Using these data, the quality coefficients of the 9 questions can be estimated using the model presented in Figure 1. This model has been called the classical MTMM model, first developed by Jöreskog (1971). The y_{ij} are the observed variables, F_i represent the traits (i.e. concepts of interest), and M_j represent the methods. The effects of the latent traits on the observed variables can be thought of as quality coefficients of the observed variables because they represent the amount of common variance in these variables that is due to the substantive latent variable of interest (in this example, satisfaction with the British government). The effects of the methods on the observed variables are called the method effects and represent the common variance among these variables that is due to the effect of the method (the response scale) rather than the latent variable of interest.

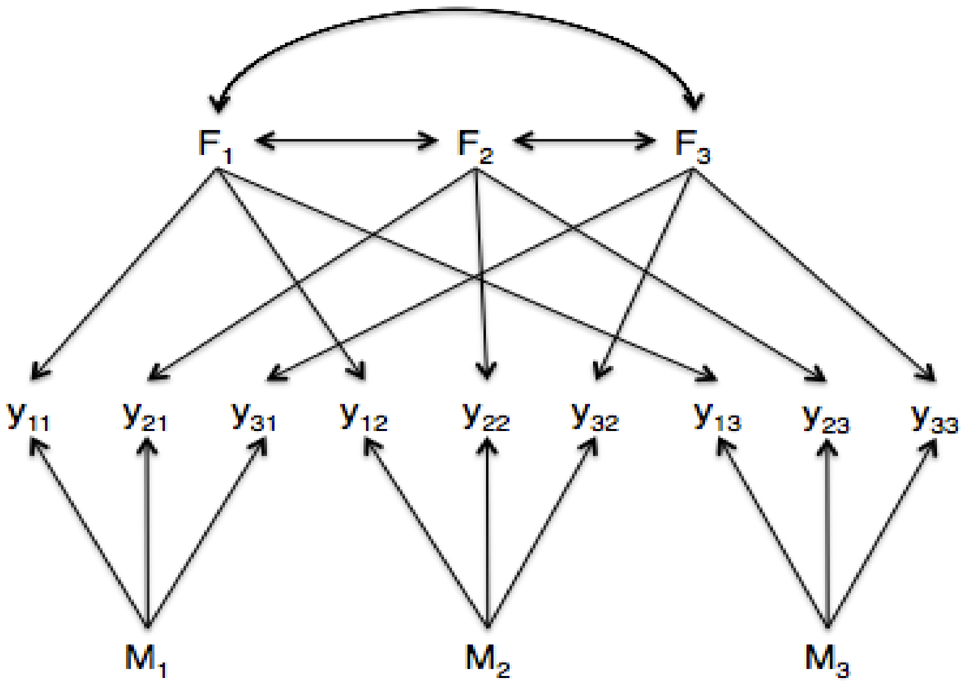

Figure 1: The classical MTMM model

The problems of the estimations of the quality and method effects are that often no estimates are obtained because of non-convergence and that one can obtain estimates with negative variances, i.e. Heywood cases, which are impossible (Rindskopf, 1984). Another problem with this approach is that one cannot separate the error components from unique components of the measures.

PHASE 2: THE IRMCS CONFERENCE

In 1989 Willem Saris organized the first meeting of the International Research group on Methodology and Comparative Survey Research (IRMCS). The purpose was to study the problems and the application of approaches to evaluate the quality of survey questions. The proceedings of this conference can be found in Saris and Van Meurs (1990). The table of content of the book is presented in Appendix 1. At the first conference three alternative designs were discussed: Quasi Simplex approach (Heise, 1969 and Alwin, 2007), the MTMM approach (Andrews, 1984) and the Repeated Multi-Method (RMM) approach (Saris, 1990). In the end, they decided that the MTMM approach promised the best results but not without several measures to minimize the problems. These decisions were:

- The use of 3 by 3 method design which has 3 traits and 3 methods.
- The questions should only differ with respect to the method not the content, i.e. no unique components.
- The method effects are expected to be equal across observed traits for the same method.
- The traits and the methods are uncorrelated.
- The methods are also uncorrelated.

These constraints have remained the same until now. Besides these basic choices, other issues were discussed and published in the book following the conference. Saris and Van Meurs carried out an experiment to determine the time needed between the observations in order to prevent memory effects. The time lag turned out to be at least 20 minutes. Költringer and Saris discussed the different steps that could be taken in order to correct misspecifications in the model if the model did not correspond with the data. Satorra looked at the problem of the use of ML estimators in the case of non-normally distributed observed variables. His conclusion was that the estimates are always consistent and the Maximum Likelihood (ML) estimator would provide proper estimates for the standard errors and the test statistic if we could assume independence from the errors of the method and the trait factors. The advantage is that in this case, one needs much smaller sample sizes than when one uses an asymptotic distribution free (ADF) estimator. The ML estimator is still used in the analyses underlying SQP. Furthermore, some remarks were made by Andrews, Mellenbergh and Molenaar regarding the way the meta-analysis of the MTMM experiments could be carried out.

After the first IRMCS conference, Andrews and Saris discussed an alternative model for the data of an MTMM experiment. They developed an equivalent model that is called the True Score Model. The alternative was developed in order to make a distinction between:

- The reliability (r^2) = 1 - random error variance.
- The validity (v^2) = 1- systematic/method effect variance.
- The quality (q^2) = reliability x validity.

A paper on this model was published by Saris and Andrews in 1991, and this MTMM model is still in use for the estimation of the quality coefficients.

PHASE 3: FURTHER IRMCS STUDIES

After the first conference, several studies were made and discussed in different conferences. These studies have been brought together in a second book edited by Saris and Münnich (1995). The table of contents of this book can be found in Appendix 2.

The chapters of Scherpenzeel, as well as De Wit and Billiet, studied the effect of unique components of questions and correlations between the methods. The unique components of questions can be avoided by using exactly the same stem of the questions in the experiments. The correlations between the methods cannot completely be avoided but if the method effects are minimal, which is often the case, then the correlations have a minimal effect on the results. If they are large they can be introduced.

Coenders et al. studied scale dependency: does it matter when we use a correlation matrix or a covariance matrix? It matters indeed. One has to analyse the covariance matrices. So we still usually use covariance matrices unless the estimation does not converge. Furthermore, Költringer detected that analysing MTMM experiments with categorical data has effects on the reliability coefficients of these variables. Therefore, the reliability of questions depends on the scale used and the size of the random errors. Coenders et al. also used in their analysis Pearson and Polychoric correlations and found that the quality estimates really differ whether one uses the one method or the other. Therefore, one cannot correct Polychoric correlations by quality estimates based on Pearson correlations and vice versa.

Together with all these studies, a large number of MTMM experiments were carried out by Andrews in the US (1984), Költringer in Austria (1993) and Scherpenzeel and Saris in the Netherlands (1997). Furthermore, all 12 countries participating in IRMCS, carried out at least one MTMM experiment. One meta-analysis was done on the basis of the data of the 12 countries (17 experiments) and another meta-analysis was done based on the data of the 5 countries (87 experiments). After these studies, several meta-analyses were done on the basis of the experiments carried out in the different countries Költringer (1993) and Scherpenzeel (1995). Additionally, a study on the categorization effects was carried out by Saris, Van Wijk and Scherpenzeel in 1998, with the following characteristics:

- 17 IRMCS experiments were carried out in 12 languages using satisfaction scales with 5, 10 and 100 scale points.

- Correlation matrices were computed using Pearson (PCC) and Polychoric (PPC) correlations. The obtained correlation matrices for 5-point scale were very different.
- The MTMM analyses showed that only the reliabilities are different across methods as expected based on the study of Költringer (1993).
- The correlations between the traits estimated with the two methods were very similar.
- The meta-analysis showed that only the reliability coefficients were different due to the differences in scales used.

The results for Pearson and Polychoric correlations for the 5-point scale were very different (Netherlands). This leads to the following conclusions with respect to the use of the Polychoric and Pearson correlations in the estimation of the quality of survey questions by MTMM experiments:

- Polychoric correlations provide information concerning continuous variables behind the observed variables.
- Pearson correlations provide information concerning the quality of the questions.
- The Pearson reliability is affected by random errors and categorization errors.
- After correction, the relationships between the latent variables are in both cases very similar.

At this point it was decided to continue with the use of the Pearson correlations.

PHASE 4: STUDIES AT THE UNIVERSITY OF AMSTERDAM

New models for MTMM data were suggested by Marsh and Browne:

- Factor model with correlated errors, the so-called Correlated Uniqueness (CU) model was suggested by Marsh (1989).
- Factor model with error correlations depending on the trait correlations. The so-called Direct Product (DP) model suggested by Browne (1984, 1985 and 1989).

Coenders and Saris (2000) showed the link between these models and the more classical MTMM models. The conclusion was that all these models could be tested using the CU model. Corten et al. (2002) did so and reported on the results of these tests:

- Data on 87 experiments of IRMCS.
- Tested on Pearson and Polychoric matrices.

The conclusions from these tests were:

- The classic or (equivalent) True Score model was in both cases the best fitting model.
- For Pearson correlations in 2% of the models, the estimation did not converge and there were no improper solutions found.
- For Polychoric correlations in 5% of the cases, the estimation did not converge and in 17% of the cases improper solutions were obtained.

In 2003 Saris and Aalberts studied different models for the explanation of correlated errors. They studied the following explanations:

1. Relative judgments
2. Style factors
3. Variation in response functions
4. Method effects

Their conclusion was that clearly the best explanation was obtained from the MTMM model.

Later Saris and Gallhofer developed a list of characteristics that should be used to explain the differences in reliability and validity for different questions. They also developed a code book for the coding of questions with respect to these characteristics.

After that, Saris and Gallhofer (2007) did a meta-analysis of experiments in 3 countries US, Austria and the Netherlands. The overview of this meta-analysis is summarized hereafter:

- There were 87 MTMM experiments done in three languages: English, German and Dutch.
- In total, there were 1,067 questions for which reliability and validity had been estimated.
- A code book was made of the characteristics of these questions.
- The relationships between the qualities and the coded question characteristics were studied through meta-analysis.
- The predictions were rather good: for reliability the R^2=0.47 and for validity R^2=0.61.

The regression equation obtained was used to make the first Survey Quality Predictor program for prediction of the quality of questions, namely the SQP (Saris, 2001). Later, this program was transformed into a Windows version, SQP 1.0, by Saris, Oberski, and Kuipers (2004). All this research was summarized in the book by Saris and Gallhofer (2007). This program has been used to make the quality predictions of European Social Survey (ESS) questions in Rounds 1 to 3.

PHASE 5: PREPARATION FOR THE ESS

The classical MTMM approach requires three repetitions of the same basic questions. Repeated observations after 20 minutes are independent, as stated by Van Meurs and Saris (1990), but they did not believe that this would also hold between the second and the third repetition. Therefore, Saris, Satorra and Coenders (2004) developed an alternative, the Split Ballot (SB) MTMM design, where each respondent is asked the same questions (observed traits) only twice in different forms or methods. This is possible by randomly splitting up the sample in three groups where the methods are distributed as presented in Table 2:

Table 2: SB-MTMM design with 3 groups

	Time 1	Time 2
Group 1	Method 1	Method 2
Group 2	Method 2	Method 3
Group 3	Method 3	Method 1

This design allows for the estimation of the reliabilities and validities of all questions but this design was not acceptable for the ESS because not all people receive the same questions in the main questionnaire. We therefore evaluated the possibility of the 2-group design. In this case, group 1 gets method 1 and 2 and group 2 gets method 1 and method 3. In this case, there are missing data by design (between the variables measured with method 2 and 3) but the reliabilities and validities could in general be estimated (Saris, Satorra and Coenders, 2004).

PHASE 6: APPLICATION IN THE ESS

The aim was to evaluate the quality of questions, compare quality across countries, and ultimately to give ESS data users the chance to correct measurement error

- In each ESS round, 4 to 6 experiments are carried out in all participating countries.
- The 2-group SB-MTMM design is used.
- Main questionnaire presents method 1.

- The Supplementary questionnaire presents method 2 and method 3.
- The True Score Model is used.
- Analyses are done with Multiple Group (MG) Maximum Likelihood (ML) estimation.

So far in the ESS, 2,700 MTMM questions have been evaluated with regard to quality, but at the same time 60,000 questions have been asked. It is impossible to evaluate all questions on the questionnaire because the repetitions increase the respondent burden. However, the already-collected information can be used for a meta-analysis which then allows the quality of questions which were not part of a MTMM experiment to be predicted.

The meta-analysis was performed by carrying out the following steps:

A list of characteristics that had been shown to affect measurement, as well as others that could potentially affect measurement, was compiled.

ESS questions for which MTMM experiments had been conducted were coded with regard to these question characteristics.

The relationships between these characteristics and the quality of the questions were estimated using the Random Forest procedure (Oberski et al., 2011).

The random forest model predicted reliability with an R^2 of 0.67 and validity with an R^2 of 0.72 using the 3,700 MTMM questions available at the time.

The prediction procedure was implemented in the program SQP 2.0.

The development of SQP 2.0 has been summarized in Saris and Gallhofer (2014), which is a revised edition of their earlier book (2007).

PHASE 7: PRESENT RESEARCH

It should be clear that the program SQP 2.0 is as good as the parts on which it is based, which are:

- The quality of the data base of experiments.
- The selection of the characteristics to make predictions.
- The quality of the estimation procedure.
- The quality of the predictions procedure.
- The quality of the program itself.

In the RECSM working paper 24 “The development of the program SQP 2.0 for the prediction of the quality of survey questions” (Saris, Oberski, Revilla, Zavala-Rojas, Lilleoja, Gallhofer and Gruner, 2011) the state of the research has been summarized. For the table of contents of the report we refer to Appendix 3.

The present research concentrates on the following issues:

7.1 New modes of survey questions

In the MTMM experiments included in SQP, the procedure was always that some questions in the main questionnaire were chosen for an experiment. Therefore, we always worked with questions asked in typical survey research. The alternative forms for these questions were chosen in such a way as to be able to form an impression of the effects of commonly used alternatives on the original questions. In this way, we tried to cover as far as possible the topics asked in survey research and the form most frequently used. However, survey research is changing fast due to the introduction of web survey and the use of tablets and smartphones. This has led to a large number of new forms of questions. Consequently, we are currently carrying out experiments with these new approaches to include them, at some point in time, in our meta-analysis and in a new version of the program SQP. For more information, refer to publications of Revilla, Saris, Loewe and Ochoa (2013), Revilla and Ochoa (2013) and Revilla (2013).

7.2 The estimation procedure for SB-MTMM models

Revilla and Saris (2013) showed that in practice, when analysing the ESS experiments, problems occur with the estimation (non-convergence and Heywood cases). In order to improve the estimation of quality, it was necessary to study under which conditions the problems occur, and what can be done to solve them. Results from Monte Carlo simulations show that the problems usually occur with the 2-group design, but not with the 3-group design. However, the 2-group design does not always have problems. Problems seem to occur when there is no correlation between the traits, when there are very similar correlations between the traits, or when some true values of the parameters are closed to a boundary. We should therefore be careful when designing the experiments in order to try to avoid these situations. Nevertheless, this is not so easy. Thus, we should also look at other ways to solve the estimation problems. Different approaches were discussed (fixing negative variances to 0; adding a third small group or even fake group; increasing the sample size; analysing all countries together). Ideas for further research on this topic were also proposed (improving Multiple Group estimation; changing the single country estimation; Bayesian estimation methods). This research on the solutions to the estimation problems is not conclusive yet and still ongoing.

7.3 The selection of characteristics for the prediction of quality

While we concentrated for a long time on the form of the questions, the experience with the proposals for questions has shown us that it is even more important to develop a procedure for creating questions that measure with a high probability the concept that a researcher wishes to measure. This means that we have paid more attention to the validity of questions, not only with regard to the effect of the method effects. Gallhofer together with Saris developed the three-step procedure for this purpose (Saris and Gallhofer, 2007). This approach has been clarified and improved recently and reported in the second edition of their book (Saris and Gallhofer, 2014).

This is also important for the design of MTMM experiments. To evaluate the quality of a question with MTMM experiments, the concepts of the questions should remain the same but the methods included in the experiments should vary. The *concept* is determined by an assertion that is created by a choice of a basic concept and the subject specific information. The method is a combination of all the choices made to transform the assertion into a question or, as Saris and Gallhofer (2007) call it, the request for an answer.

All questions are supposed to measure a concept. However, there are many types of concepts in social sciences. We can distinguish concepts-by-postulation (CP) from concepts-by-intuition (CI). CI are simple concepts, which are immediately obvious and can be measured by a single question. In contrast, CP, also called constructs, require an explicit definition. They can therefore not be measured by a single question but are based on several CI

The number of possible specific concepts[1] is infinite. However, most of these specific concepts can be classified in a reduced number of basic concepts[2]. For a link between specific concepts and basic concepts we refer to page 32 in Saris and Gallhofer (2014) where many examples are given.

[1] Specific concepts are for example: political trust, happiness, media use etc.

[2] In previous presentations, as well as in Saris and Gallhofer (2007), the 'basic concepts' were labeled 'abstract concepts' and embrace evaluative belief, feeling, importance (of something), expectation of future events, facts, background, or behaviour, and others.

There are so many different questions possible for one CI that it is not possible to specify a question from a basic concept but it is possible to transform a basic concept into an assertion. There are only three basic forms of assertions:

1. Subject + Predicator + Subject Complement – e.g. 'Clinton was a good president'.
2. Subject + Predicator + Direct object – e.g. 'The President likes his job'.
3. Subject + Predicator – e.g. 'The president has changed'.

When an assertion coming from the basic concept is defined, it can be formulated for a specific concept by adding the subject specific information. Therefore, specific concepts are defined by a basic concept plus the subject specific information, called 'domain' by Saris and Gallhofer (2007) and in SQP. This defines the stem of a question and for the MTMM experiments the stem of the question has to be kept the same in order to ensure that the concept remains the same.

We call a 'method' all steps taken to transform an assertion into a request for an answer. The details of this approach are discussed in Saris and Gallhofer (2014)

7.4 The prediction model of SQP 2.0

The prediction model was estimated using random forests of regression trees. The Random Forests Algorithm (Breiman, 2001) can be summarized as:

- Take a random sample without replacement of the cases or questions: some of them will be considered 'in-bag' cases and the rest will be 'out-of-bag' cases.
- Take a sample of the characteristics.
- Grow a regression tree on the 'in-bag cases' using the subset of selected features.
- Calculate mean square prediction error on the 'out-of-bag' cases and build a cross-validation.
- Back to first point until 1,500 trees have been grown.
- Missing cases will be imputed using Multiple Imputation with chained equations (van Buuren and Groothuis-Oudshoorn, 2011) and a small random forest is grown for each imputed dataset. These are joined to form the final forest.

The advantage of this procedure is that the predictions are better than the predictions using SQP 1.0 however the Random Forest approach has the disadvantage that it gives predictions but does not provide clear information regarding the importance of the different factors which determine the quality of the questions. This is the advantage of the previous used regression analysis. Therefore, a further analysis of a regression analysis of the new rather extended data set is planned.

7.5 The program SQP 2.0

While the program SQP 2.0 functions well, there are always reasons to improve such programs. Therefore, we are currently working on improvements. Besides that, we have added information regarding limitations of the program, frequently asked questions, a general introduction, a coding guideline, the acknowledgement of the work of researchers in the past, the publications about and using SQP and here we have added an overview of the basic studies on which SQP is based.

APPENDICES

Appendix 1

Book title: Evaluation of measurement Instruments by meta-analysis of Multitrait Multimethod studies

Authors: W.E. Saris and A. van Meurs, editors

Year: 1990

Table of contents:

Appendix 2

Book title: The Multitrait-Multimethod Approach to Evaluate Measurement Instruments

Authors: W.E. Saris and Á. Münnich

Year: 1995

Table of contents:

10. A modified True score MTMM model for analyzing stereotype effecs of characterizations of nations

György Hunyady and Ákos Münnich

Part II: Meta analysis and applications

11. Design of Meta Analysis

Annette Scherpenzeel

12. Measurement Quality in Austrian Personal Interview Surveys

Richard Költringer

13. Meta Analysis of European comparative study

Annette Scherpenzeel

14. Correction for Measurement error in life satisfaction research

Willem E. Saris and Annette Scherpenzeel

15. Correction for measurement error in satisfaction with life study

Dagmar Krebs and Peter Schmidt

Appendix 3

Book title:	The development of the program SQP 2.0 for the prediction of the quality of survey questions
Authors:	W. E. Saris. D. L. Oberski, M. Revilla, D. Zavala-Rojas, L. Lilleoja, I. N. Gallhofer and T. Gruner
Year:	2011

Table of contents:

REFERENCES[3]

Alwin, D. F. (2007). Margins Errors. A study of reliability in survey measurement. *Hoboken: Wiley.*

Andrews, F. M. (1984). Construct Validity and error components of survey measures: structural equation approach. *Public Opinion Quarterly, 48, pp. 409-442.*

Breiman, L. (2001). Random forests. *Machine learning, 45(1), 5–32.*

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

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

Browne, M. W. (1989). Relationships between an additive model and a multiplicative model for multitrait-multimethod matrices. *In R. Coppi and S. Bolasco (Eds.), Multiway data analysis (pp. 507-250), Amsterdam: North-Holland.*

Buuren, S., and Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. *Retrieved from http://doc.utwente.nl/78938/*

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

Coenders, G., Saris, W. E., Batista-Foguet, K. M. and Andreenkova, A. (1999). Stability if three-wave simplex estimates of reliability. *Structural Equation Modeling, 7, pp. 219-250*.

Coenders, G. and Saris, W. E. (2000). Testing Nested Additive, Multiplicative, and General Multitrait-Multimethod Models. *Structural Equation Modeling: A Multidisciplinary Journal, 7(2), pp 219-250. Available at: http://www3.udg.edu/fcee/professors/gcoenders/pap4.pdf*

Corten, I. W., Saris, W. E., van der Veld, W. M., Aalberts, C. E., and Kornelis, C. (2002). Fit of Different Models for Multitrait-Multimethod Experiments. *Structural Equation Modeling: A Multidisciplinary Journal.doi:10.1207/S15328007SEM0902.*

Heise, D. R. (1969). Separating reliability and stability in test-retest-correlation. *American Sociological Review, 34,93-101.*

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36, 109–133. doi:10.1007/BF02291393.*

Költringer, R. (1993). Gültigkeit von Umfragedaten. *Wien: Böhlau.*

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13, 335–361.*

Oberski, D. L. (2011). *Measurement error in comparative surveys* (Doctoral dissertation, Universiteit van Tilburg).

Oberski, D., Gruner, T. and Saris, W.E. (2011). The program SQP 2.0 for prediction of quality of questions and its applications. *In Saris et al. (2011) The development of the program SQP 2.0 for the prediction of the quality of survey questions. RECSM Working Paper 24. Available at: http://www.upf.edu/survey/_pdf/RECSM_wp024.pdf*

Revilla, M. (2013). Effect of using different labels for the scales in web surveys. *RECSM working paper 30, to be published in International Journal of Market Research 2015.*

Revilla, M. and Saris, W. E. (2013). The Split-ballot Multitrait-Multimethod Approach: Implementation and Problems. *Structural Equation Modeling: A Multidisciplinary Journal, 20(1), pp. 27-46.*

Revilla, M. and Ochoa, C. (2013). Quality of different scales in an online survey in Mexico and Columbia. *RECSM working paper 34.*

[3] If the publications cannot be obtained directly, a request can be send to sqp@upf.edu.

Revilla, M, Saris, W. E., Loewe, G. and Ochoa, C. (2013). Can an online panel oriented to marketing surveys get similar SEM estimates of question quality as the face-to -face European Social Survey. *RECSM working paper 33.*

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods and Research 13, 109–119.*

Saris, W. E. and van Meurs, A. (Eds.) (1990). Evaluation of measurement instruments by meta-analysis of multitrait multimethod studies. *Amsterdam: North Holland.*

Saris, W. E. and Andrews, F. M. (1991). Evaluation of measurement instruments using a structural modeling approach. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. Mathiowetz, and S. Sudman (eds.), *Measurement Errors in Surveys*. *New York: Wiley, pp. 575-559.*

Saris, W. E. and Münnich, Á. (1995*).* The multitrait-multimethod approach to evaluate measurement instruments. *Budapest: Eötvös University Press.*

Saris, W. E., Van Wijk, T. and Scherpenzeel, A. (1998). Validity and reliability of subjective social indicators. *Social Indicators Research, 45, 173–199.*

Saris W. E., and Aalberts, C. (2003). Different explanations for correlated errors in MTMM studies. *Structural modeling*, *10, pp. 193-214.*

Saris, W. E. (2001). SQP: survey quality predictor. *DOS application program.*

Saris, W. E., Oberski, D. L. and Kuipers, S. (2004). SQP*:* survey quality predictor. *Windows application program.*

Saris, W. E., Satorra, A., and Coenders, G. (2004). A New Approach to Evaluating the Quality of Measurement Instruments: The Split Ballot MTMM Design. *Sociological Methodology*, *34(1), 311–347. doi:10.1111/j.0081-1750.2004.00155.*

Saris, W. E. and Gallhofer, I. N. (2007). Design, Evaluation, and Analysis of Questionnaires for Survey Research. *Hoboken, NJ, USA: John Wiley & Sons, Inc.*

Saris, W. E., Oberski, D., Revilla, M., Zavala-Rojas, D., Lilleoja, L., Gallhofer, I. N., and Gruner, T. (2011). The development of the program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working Paper 24. Available at:* *http://www.upf.edu/survey/_pdf/RECSM_wp024.pdf*

Saris, W.E. and Gallhofer, I.N. (2014). Design, Evaluation and Analysis of Questionnaires for Survey Research. (Second edition). *New York: Wiley.*

Scherpenzel, A. C. (1995). A question of quality: Evaluating survey questions by Multitrait-Multimethod studies. *Leidschendam, KPN.*

Scherpenzeel, A. C., and Saris, W. E. (1997). The validity and reliability of survey questions. *Sociological Methods & Research*, *25*(3), *341–374. doi:10.1177/0049124197025003004*

Zavala-Rojas, D. and Saris, W. E. (2013). A procedure to prevent differences in translated survey items using SQP. *RECSM Working Paper 3*8. Available at: http://www.upf.edu/survey/_pdf/RECSM_wp038.pdf