Correction for Measurement Errors in Survey Research: Necessary and Possible

Willem E. Saris · Melanie Revilla

Accepted: 1 June 2015 / Published online: 17 June 2015
© Springer Science+Business Media Dordrecht 2015

Abstract  Survey research is the most frequently used data collection method in many disciplines. Nearly, everybody agrees that such data contain serious measurement errors. However, only few researchers try to correct for them. If the measurement errors in the variables vary, the comparison of the sizes of effects of these variables on each other will be wrong. If the sizes of the measurement errors are different across countries, cross-national comparisons of relationships between variables cannot be made. There is ample evidence for these differences in measurement errors across variables, methods and countries (Saris and Gallhofer in Design, evaluation and analysis of questionnaires for survey. Wiley, Hoboken, 2007; Oberski in Measurement errors in comparative surveys. PhD thesis, University of Tilburg, 2011). Therefore, correction for measurement errors is essential for the social sciences. The correction for measurement errors can be made in a simple way, but it requires that the sizes of the error variances are known for all observed variables. Many experiments are carried out to determine the quality of questions. The relationship between the quality and the characteristics of the questions has been studied. Because this relationship is rather strong, one can also predict the quality of new questions. A program SQP has been developed to predict the quality of questions. Using this program, the quality of the questions (complement of error variance) can be obtained for nearly all questions measuring subjective concepts. For objective variables, other research needs to be used (e.g., Alwin in Margins of error: a study of reliability in survey measurement. Wiley, Hoboken, 2007). Using these two sources of information, making correction for measurement error in survey research is possible. We illustrate here that correction for measurement errors can and should be performed.

Keywords  Correction for measurement errors · Quality · SQP 2.0 · European Social Survey
1 Introduction

Most studies that require information of individuals about values, attitudes, opinions, evaluations, feelings, preferences, expectations, status, occupation, education, income and behavior rely on interviews or questionnaires. As a consequence, survey research is the most frequently used method for collecting data in Sociology, Political Science, Communication Science and Marketing Research (Saris and Gallhofer 2007).

The effects that the wording of survey questions can have on their responses have been studied by many researchers, to mention some important contributions: Belson (1981), Schuman and Presser (1981), Sudman and Bradburn (1982), Andrews (1984), Alwin and Krosnick (1991), Molenaar (1986), Költringer (1993), Scherpenzeel (1995), Tourangeau et al (2000), Dillman (2000), Alwin (2007), Saris and Gallhofer (2007) and Biemer (2011). In all these studies, the researchers indicate that the formulation of the questions has a considerable effect on the results one obtains. That is the same as saying that there is a considerable error in survey measurement, even though in many cases we do not know what the true values of the variables we want to measure are.

While these studies are very well known to the research community and it is a very common opinion that survey data contain a lot of measurement errors, only very few researchers try to correct for these errors. To illustrate this point, we have collected information for a number of important journals with respect to the frequency of use of survey research, the attention paid to measurement errors and the correction for these errors. Table 1 summarizes these results.

This table shows not only how important survey research is in the chosen journals but also how little attention has been paid to measurement problems in these journals, let alone that correction of the measurement errors has been performed.¹

One may wonder how this lack of attention to measurement errors can go together with the general idea that survey research contains a lot of errors as has been shown by the above-mentioned studies. We can see three main possible explanations:

1. The effect of the measurement errors and their consequences are relatively small, so they can be ignored.
2. The procedures to correct for measurement errors are so complex or expensive that in most research these corrections cannot be performed.
3. The estimates of the measurement error variances or the complement of that, the quality of questions, are not available, and so correction is not possible.

In this paper, we would like to discuss these three possible explanations. We want to show that the effects of measurement errors are considerable and cannot be ignored, that corrections can be made easily and that nowadays estimates of the size of the errors variances or the quality of questions are available. As a consequence, we think that all researchers can, but also should, correct for measurement errors in order to provide believable results of their research. This is what we want to show in this paper. We will discuss the three issues in sequence and, then, come back to the general conclusions.

¹ We are grateful to Wiebke Weber of RECSM for collecting the data in Table 1. Pei-Shan Liao of the Academia Sinica and Zih-Wei Wang of the National Taipei University reported to us that in the Taiwanese Sociological Journal and the Political science review, the situation was very similar: Together these journals published in 2010–2012 in total 67 papers of which 18 used survey research and 9 mentioned the problem of measurement errors but only one makes corrections for measurement errors. We are very grateful for this information.
2 Can measurement Errors in Survey Research be Ignored?

In several studies (Andrews 1984, Alwin and Krosnick 1991, Költzing 1995, Scherpenzeel 1995 and Saris and Gallhofer 2007), it has been shown that the measurement errors in survey questions are considerable. Alwin (2007) suggests that 50% of the variance of the observed variables in survey research is due to errors. So there is a considerable difference between the variable one likes to measure and the one that is really measured with the question. This difference has a big effect on conclusions of research. It is a fundamental problem of these sciences, as we will now demonstrate.

Imagine that we would like to know the strength of relationship between two opinions (the variables of interest); for example, job satisfaction ($f_1$) and life satisfaction ($f_2$) represented by the correlation coefficient $q(f_1, f_2)$. This coefficient cannot be obtained directly. One can only estimate the relationship $q(y_1, y_2)$ between the observed variables, i.e., the responses to questions with respect to job satisfaction ($y_1$) and life satisfaction ($y_2$). The relationship between $f_1$ and $y_1$ and between $f_2$ and $y_2$ will not be perfect because of the measurement errors ($e_1$ and $e_2$). The standardized effect of the variable of interest $f_i$ on $y_i$ is called the quality coefficient ($q_i$). This simple idea is presented in Fig. 1. A more elaborate measurement model will be presented later.

If the variables of interest (opinions) and the errors are uncorrelated and the variables are standardized, the variance of the observed variable is 1 and it follows that:

$$\text{var}(y_i) = 1 = q_i^2 + \text{var}(e_i)$$

Because of this result, the quality of the $i$th question is $q_i^2 = 1 - \text{var(errors)}$ and the coefficient $q_i$ is called the quality coefficient.

It can also be shown, using path analysis, that the following relationship exists between the correlation for the observed answers $\rho(y_1, y_2)$ and the correlation for the latent variables of interest $\rho(f_1, f_2)$:

$$\rho(y_1, y_2) = \rho(f_1, f_2)q_1q_2$$

Table 1  Attention to measurement problems in social science journals of 2011

| Journal | Year | No. paper | Survey research used | Errors mentioned | Errors corrected |
|---------|------|-----------|----------------------|------------------|-----------------|
| AJPS    | 2011 | 54        | 23                   | 3                | 0               |
| EJPR    | 2011 | 32        | 20                   | 4                | 1               |
| ESR     | 2011 | 48        | 41                   | 9                | 1               |
| JM      | 2011 | 47        | 27                   | 11               | 8               |
| POQ     | 2011 | 33        | 32                   | 4                | 1               |
| SOCI    | 2011 | 48        | 41                   | 13               | 6               |

APS R The American Journal of Political Science, EJPR European Journal of Political Research, ESR The European Sociological Review, JM Journal of Marketing, POQ Public Opinion Quarterly, SOCI Social Indicators Research (in this journal, there were 150 papers published in 2011. Therefore, a systematic sample of 48 papers has been drawn)
The two correlations are only equal if the quality of both measures is perfect (equal to 1); i.e., there are no measurement errors. Unfortunately, this will never occur. What happens if the qualities of the measures are different from 1.0 is presented in Table 2. In the example, we assume for illustrative purpose that the correlation is .9 between the two variables of interest so $\rho(f_1, f_2) = .9$. Whenever the quality of the two variables goes down, $\rho(y_1, y_2)$ will also go down but much faster. If the quality of the measures is equal to .5 (the average quality in survey research, Alwin 2007), then the quality coefficients $q_i$ are .7 and the expected correlation between the observed variables will be only half of the size of the correlation between the variables of interest. If the quality coefficients go down to .6, then this correlation will be as small as a third of the true value. The correlations between the variables of interest are very much underestimated, if one does not correct for measurement errors.

However, this is not the only problem. Measurement errors also make comparisons of correlations questionable. Imagine that a researcher is interested in the correlation of age and job satisfaction with life satisfaction. Imagine that the effect of age on life satisfaction is .4 and the effect of job satisfaction is .6. The quality of the measurement of age will be nearly perfect. But we can expect quite some measurement errors for the two other variables. Let us assume that they both have a quality coefficient of .6. In Fig. 2, we have presented this model.

The correlation between the observed variables LS and age will be .238 ($=.4 \times .99 \times .6$), and the one between LS and job satisfaction will be .22 ($=.6 \times .6 \times .6$). Apart from the fact that, due to errors, these correlations are much lower than the correlations between the variables of interest, the researchers may also draw the wrong conclusion that age is a bit more correlated than job satisfaction with life satisfaction. The true correlation between life satisfaction and job satisfaction is much larger (.6) than the correlation between life satisfaction and age (.4).

For the same reasons, comparisons of relationships across countries or cultural groups cannot be made if one does not know whether the measurement errors are comparable. To illustrate this point, imagine that in two countries the correlations between two latent variables are equal to .8, but the quality of the measures is rather different, say for both measures $q = .9$ in country A and $q = .7$ in country B. Then, using Eq. 2, one can derive that the correlations between the observed variables will be .65 in country A and .40 in country B.

Overall, the conclusion should be that without correction for measurement errors, one runs the risk of very wrong conclusions with respect to relationships between variables and differences in relationships across countries.
3 Is it Difficult to Correct for Measurement Errors?

The commonly known procedure to correct for measurement errors using structural equation models (SEMs) has already been introduced in Goldberger and Duncan (1973) together with the LISREL program for estimation of these models with latent variables (Jöreskog 1973). A “simple” example of such a model is presented in Fig. 3. Here the researcher has made a model for explanation of “Environmental friendly behavior” using two endogenous variables “Environmental values” and “Influence” and two exogenous variables “Perception of environmental damage” and “understanding politics.” All these variables are latent variables. Each of these latent variables is measured by two indicators. This can be single questions or two composite scores based on several indicators.

Essential in this approach to correct for measurement errors is that one needs for each latent variable at least two observable indicators. Otherwise, the error variances cannot be estimated and corrected. This is also the major problem of this approach. The research costs increase. This approach also increases the length of the survey, the burden for the respondents and the complexity of the models. As a consequence, the estimation and testing of these models become more difficult as well. All these reasons may have been enough to prevent the use of this in principle correct procedure (see Table 1).

But correction for measurement errors does not have to be so difficult. A much simpler procedure to correct the correlation matrix or covariance matrix for measurement errors has been developed earlier. After that, one can analyze the data as if there were no measurement errors. This approach follows directly from Eq. 2:
Equation 3 suggests that the correlation between the variables of interest is equal to the correlation between the observed variables divided by the product of quality coefficients of the measures used. Thus, correction for measurement error in the observed correlation is very simple if the quality estimates of the observed variables are known. This result holds for single questions as well as composite scores. This result was already known in psychology for a long time (Lord and Novick 1968). However, also this approach is hardly used.

Let us illustrate this two-step procedure with an example from a recent study of opinions about democracy in Europe. The data have been collected in the pilot study of the 6th round of the European Social Survey (ESS). Using Mokken scaling (Mokken 1971), the composite scores on two latent variables were obtained:

- The first one, based on opinions about liberal rights, was called “liberal democracy,”
- The second one, based on opinions about electoral requirements, was called “electoral democracy.”

The Mokken scale also provides an estimate of the quality of the measure developed. The qualities of these two scales turned out to be .79 for liberal democracy ($q_1^2$) and .77 for electoral democracy ($q_2^2$), while the estimated correlation $\rho(y_1, y_2)$ between the two scales was .638. To correct for measurement errors in this correlation, we use Eq. 3 and we get:

$$\rho(f_1, f_2) = \frac{\rho(y_1, y_2)}{q_1 q_2}$$
We see that in this case the correlation increases of about 20 % when correcting for measurement error.

In the study mentioned above, it is also expected that the variable “liberal democracy” should correlate with the variables measuring opinions about the importance for the democracy of preventing poverty, holding referenda and sufficiently high incomes for the people. For these observed variables, the quality has also been estimated using SQP 2.0. For the opinion about poverty, called “Just,” the quality was .51, for the opinion about referenda, called “Direct,” the quality was .62, and for the household income, called “Income,” the quality was .92.

As mentioned above, correction for measurement errors in further analysis of data is rather simple after correction of the correlations or covariances for measurement errors. The analysis with correction for measurement errors can be implemented in a simple way with programs which can use as data the covariance or correlations matrix like LISREL, STATA or the Lavaan module for R. For more information and examples, we refer to the ESS training modules (De Castellarnau and Saris 2014).3

Appendix 1 presents the input for the analysis with and without correction for measurement errors using the program LISREL. The only difference between the inputs without and with correction for measurement errors is the diagonal of the correlation matrix. Without correction for measurement errors, the original correlations are used, whereas with correction for measurement errors, the diagonal of the matrix contains the estimates of quality of the measures. This is the same as the original variance, which is (in this case) 1 minus the estimated error variance. This change is enough to get the estimates of the regression coefficients corrected for measurement errors.

The program computes the correlation matrices correcting for measurement errors. This is done using the formula mentioned in Eq. 3. It leads to considerable changes in the correlations between the variables, as can be seen in Table 3. As a consequence, the estimated effects of the different variables also considerably change when correcting for measurement errors.

While without correction for measurement errors, all three variables have significant effects on the opinion about liberal democracy, after correction, the effect of the variable “Just” is nearly twice as large and the effect of the variable “Direct” is a fourth of what it was before and is even not significant anymore.

To conclude, since correction for measurement errors is simple, it can and must be done because the results can be very different with and without correction for measurement errors. However, it is necessary that good estimates of the qualities of the questions are available.

4 Are Estimates of the Quality of Survey Questions Missing?

4.1 Estimation of the Quality of Questions

There are a lot of different procedures to estimate the quality of questions and measures for complex concepts. Maybe the most well known is the test–retest design (Lord and Novick 1968) to estimate the reliability of questions. An adjustment of this approach is the quasi-simplex model (Heise 1969; Wiley and Wiley 1970) used by Alwin and Krosnick (1991)
and (Alwin 2007). The multitrait-multimethod (MTMM) design was suggested by Campbell and Fiske (1959) to take the effects of the method used into account. For concepts with multiple indicators, different procedures have been developed based on latent variable models like factor analysis (Lawley and Maxwell 1971; Harman 1976) and latent class analysis (Hagenaars 1988; Vermunt 2003; Breiman 2001). Besides that, scaling methods have been developed, like Thurstone scale, Likert scale (Torgerson 1958), Guttman scale and Mokken scale (Mokken 1971), unfolding scale (Van Schuur 1997), Rasch scale (Rasch 1960) and item response theory (Hambleto et al. 1991). For the advantages and disadvantages of these different procedures, we refer to this literature.

All these procedures require at least two questions for each concept. That means that the number of questions is at least twice the number of concepts one likes to evaluate with respect to quality. As a consequence, these procedures lead to rather costly and time-consuming research with rather complex procedures. Besides, all these procedures provide the estimates of the quality of questions specific to the formulations of the questions used in the specific questionnaire and context. Generalization is not easily possible.

Therefore, before the final data collection, a lot of research has to be carried out in order to determine the quality of all variables to be used. This is so much work that it is only seldom done, as can be seen in Table 1. Thus, the question is whether there is a procedure that is less time- and money-consuming to estimate the quality of survey questions.

### 4.2 Prediction Instead of Estimation

The alternative to estimate the quality of questions is to predict the quality of the questions. This can be done if one has a sufficiently large database of questions for which the qualities are known and one has found an algorithm that can predict with high precision the quality of these questions on the basis of their characteristics and the context in which they have been asked. If this is the case, one can also apply the algorithm on new questions to predict their quality and one does not have to collect any extra data to estimate the quality of these new questions.

First, this approach has been worked out in three countries and led to the program SQP 1.0 (Saris and Gallhofer 2007). Later, it was extended to many different European

| Table 3 Correlations and regressions from LISREL |
|-----------------------------------------------|
| Without correction for measurement errors     | With correction for measurement errors |
| Correlations                                   | Correlations                                  |
|                                               | Liberal | Just | Direct | Income |
| Liberal                                      | 1.00    |      |        |        |
| Just                                         | .50     | 1.00 |        |        |
| Direct                                       | .40     | .41  | 1.00   |        |
| Income                                       | .21     | -.05 | -.12   | 1.00   |
| Regression (36 % explained)                   |         |      |        |        |
| Liberal                                      | .40     | .27  | .26    |        |
| s.e.                                          | (.05)   | (.05)| (.04)  |        |
| t value                                       | 8.77    | 5.84 | 6.29   |        |
|                                               | Liberal | Just | Direct | Income |
|                                               | .76     | .07  | .32    |        |
| s.e.                                          | (.04)   | (.04)| (.03)  |        |
| t value                                       | 18.22   | 1.59 | 11.06  |        |

and (Alwin 2007). The multitrait-multimethod (MTMM) design was suggested by Campbell and Fiske (1959) to take the effects of the method used into account. For concepts with multiple indicators, different procedures have been developed based on latent variable models like factor analysis (Lawley and Maxwell 1971; Harman 1976) and latent class analysis (Hagenaars 1988; Vermunt 2003; Breiman 2001). Besides that, scaling methods have been developed, like Thurstone scale, Likert scale (Torgerson 1958), Guttman scale and Mokken scale (Mokken 1971), unfolding scale (Van Schuur 1997), Rasch scale (Rasch 1960) and item response theory (Hambleto et al. 1991). For the advantages and disadvantages of these different procedures, we refer to this literature.

All these procedures require at least two questions for each concept. That means that the number of questions is at least twice the number of concepts one likes to evaluate with respect to quality. As a consequence, these procedures lead to rather costly and time-consuming research with rather complex procedures. Besides, all these procedures provide the estimates of the quality of questions specific to the formulations of the questions used in the specific questionnaire and context. Generalization is not easily possible.

Therefore, before the final data collection, a lot of research has to be carried out in order to determine the quality of all variables to be used. This is so much work that it is only seldom done, as can be seen in Table 1. Thus, the question is whether there is a procedure that is less time- and money-consuming to estimate the quality of survey questions.

### 4.2 Prediction Instead of Estimation

The alternative to estimate the quality of questions is to predict the quality of the questions. This can be done if one has a sufficiently large database of questions for which the qualities are known and one has found an algorithm that can predict with high precision the quality of these questions on the basis of their characteristics and the context in which they have been asked. If this is the case, one can also apply the algorithm on new questions to predict their quality and one does not have to collect any extra data to estimate the quality of these new questions.

First, this approach has been worked out in three countries and led to the program SQP 1.0 (Saris and Gallhofer 2007). Later, it was extended to many different European
countries participating in the ESS and led to the development of the program SQP 2.0 (Oberski et al. 2011; Saris and Gallhofer 2014) which is free of charge and available for use at sqp.upf.edu.

Since 2002, each round of the ESS contains 4–6 experiments to evaluate the quality of the questions, carried out in most countries. Consequently, after 6 rounds, the quality is now known\(^4\) for more than 5000 questions in more than 20 different languages.

Concerning the way the program SQP 2.0 has been developed, we refer to Saris et al. (2011) and Saris and Gallhofer (2014). Below, we give only a brief description of the program.

The quality of questions was determined analyzing the data of MTMM experiments (Andrews 1984; Saris et al. 2004). In this way, we obtained estimates of the reliability (1 – random errors variance), validity (1 – method error variance) and the quality (product of reliability and validity) of all questions involved in the experiments and in all the different countries.

The idea was to use the characteristics and the context of the questions as predictors of the quality of questions. Therefore, we have made a program to code the questions that were involved in the experiments of the ESS. For details of these characteristics, we refer to Saris et al. (2011). People who were native speakers in the different languages involved in the ESS and were able to understand English coded the questions. Results were controlled by comparing the codes from different languages with the codes of the English source questionnaires, and corrections were made in the codes when necessary.

The next step was to choose a procedure to study the relationship between the question characteristics and the quality estimates of these questions. For this purpose, we have not chosen the regression model used in the past (Saris and Gallhofer 2007) but the so-called “Random forest” approach developed by Breiman (2001), because it was suggested to be the most efficient prediction procedure for this kind of problems.

It turned out that this procedure indeed provided better predictions of the reliability and validity for our data. The \(R^2\) was .69 for reliability and .72 for validity.

Finally, based on this knowledge, this algorithm has been used to develop the computer program SQP 2.0\(^5\) to generate predictions of the quality of questions (Oberski et al. 2011).

In order to predict the quality of new questions, the user has to code the characteristics of the question. Then, the program generates the prediction of the quality of the question. This means that researchers can now get, via SQP, a prediction of the quality for most ESS questions but also for other new questions without further costs than the time required to introduce the question in the program and to code it.

So far, we discussed the estimation and prediction of the quality of single questions. Often, researchers use concepts based on several indicators. Therefore, we need also a solution for the prediction of the quality of composite scores for complex concepts. Such a solution exists based on the predictions of the quality of single questions. For this topic, we refer to Saris and Gallhofer (2014, chapter 14) and the ESS online training module (De Castellarnau and Saris 2014) mentioned above.

We can conclude that, nowadays, a simple procedure is available to obtain the quality of existing and new questions and composite scores for concepts with multiple indicators.

---

\(^4\) However, this information is not enough to correct for measurement errors in all ESS questions because at the same time in the ESS more than 60,000 questions were asked with respect to values, opinions, attitudes, preferences, feelings, etc. This illustrates that a different approach is required.

\(^5\) SQP is free of charge available at sqp.upf.edu.
5 A realistic Illustration

As an illustration of this approach, we have chosen a research issue that has been studied by many researchers recently on the basis of the ESS data of Round 3. It is the explanation of opinions of people about extra immigration in their country. Some of the variables introduced in the ESS for explanation of this opinion and the model proposed are presented in Fig. 4.

The questions asked to measure these concepts are presented in Appendix 2. In this study, the questions B37 till B40 have been used, respectively, for the variables “Allow,” “Economic threat,” “Cultural threat” and “Better live.” The question “Allow” is quite different from the other three which have been specified with the same type of scale. This can lead to a standard reaction of the respondents which is called the method effect (Campbell and Fiske 1959), for example giving each time the same answer (straightlining) or using the scale in a specific way (response style). Because this effect will occur in all three observed variables, one can expect an extra correlation between these variables. This correlation is called the common method variance (CMV). This means that our earlier simple measurement model should be adjusted to take the method effect into account, as presented in Fig. 5.

Equations 2 and 3 have also to be changed in order to take this CMV into account. The more realistic equations are presented in Eqs. 4 and 5.

\[ \rho(y_{1j}, y_{2j}) = \rho(f_1, f_2)q_{1j}q_{2j} + \text{CMV} \]  

(4)

From which follows as before:

\[ \rho(f_1, f_2) = \left[ \rho(y_{1j}, y_{2j}) - \text{CMV} \right] / q_{1j}q_{2j} \]  

(5)

The predicted quality \( q^2 \) of the questions decomposed into reliability \( r^2 \), validity \( v^2 \) and method effect \( m^2 \) is presented in Table 4.

In the top part of Table 5, the correlations between the observed variables have been presented below the diagonal. Above the diagonal, the CMV\(^6\) for the different correlations is presented. In the lower part of the table, the CMV is subtracted from the observed correlations, and on the diagonal, the quality of the different measures are placed as before. This latter covariance matrix has been used for the estimation of the effects in the model corrected for measurement errors. The correlation matrix below the diagonal at the top part of the table is used for the estimation of the effects without correction for measurement errors.

\(^6\) It can be shown that the CMV = \( r_{ij}m_{ij}m_{2j}r_{2j} \).
The results of these two analyses are presented in Table 6. The table shows again the considerable differences in the effects depending whether one corrects for measurement error or not.

While the economic threat has a significant effect in both regression equations if one does not correct for measurement errors, this effect is reduced to zero after correction for measurement errors, while the effects of Better life and Cultural threat become much larger. Also, the explained variance is increased considerably as expected when the random error is removed from the variance of the dependent variables.
This result shows once again the importance of correction for measurement errors in the analyses.

6 Conclusions

In this paper, we have discussed three possible reasons why researchers in the social sciences hardly correct for measurement errors.

The first reason was that the effect of these errors may be ignorable because they are very small. Based on a theoretical argument, empirical research and several illustrations, we have shown that this is not the case. The effects can be considerable. If one does not correct for measurement errors, the consequences are the following:

1. The estimates of the relationships between variables are in general biased: They can be underestimated because the quality is below 1 or overestimated because of high method effects.
2. Thus, the estimates of effects of different variables can lead to wrong conclusions.
3. The estimates of the relationships across countries cannot directly be compared.

Thus, correction for measurement errors is absolutely necessary.

The second possible reason for not correcting for measurement error was that the correction procedures are too complicated. However, we showed that a very simple procedure is available for correction. We showed that any model can be estimated correcting for measurement errors by adjusting the covariances for CMV and subtracting from the variances of the variables in the covariance matrix the error variance of each variable. In this case, one gets automatically estimates of the parameters in the model corrected for measurement errors. All steps can be done in a SEM program like LISREL, but it can also be done in other programs. The general conclusion is that there are simple procedures for correction for measurement errors available. So this cannot be a reason to ignore the presence of measurement errors.

Table 6  Estimates of the parameters with and without correction

|                     | Without correction On allow immigration | With correction for errors On allow immigration |
|---------------------|----------------------------------------|-----------------------------------------------|
| By                  |                                        |                                               |
| Better life         | -.265*                                 | -.609*                                        |
| Economic threat     | -.133*                                 | .001                                          |
| Cultural threat     | -.154*                                 | -.140*                                        |
| Total explained ($R^2$) | .254                                   | .547                                          |
|                     | On better life                         | On better life                                |
| By                  |                                        |                                               |
| Economic threat     | -.310*                                 | -.007                                         |
| Cultural threat     | .500*                                  | .938*                                         |
| Total explained ($R^2$) | .564                                   | .868                                          |

* Indicates that the coefficient in absolute value is more than twice as large as the standard error of this coefficient
The third possible reason was that the size of the measurement error variance or the quality of the measures is not available. It is indeed quite some work to collect information of the quality of all variables of interest. However, we presented in this paper a new approach based on the meta-analysis of a lot of measurement quality experiments which makes it possible to predict the quality of nearly all substantive variables just by coding the characteristics of the question of interest. This procedure is available in the program SQP 2.0 which is freely available for use.

Given this situation, correction for measurement errors can routinely be done. There is no reason anymore to ignore measurement errors.

Above, we mentioned that an alternative for the procedure suggested here is the multiple indicators approach. This approach can be applied in combination of latent variable models (as illustrated before) or with models without latent variables but with composite scores as observed variables. Both approaches are in principle correct, but they seem too complex for researchers because Table 1 shows that they are only seldom used. The problem of the multiple indicators models combined with latent variables is that the models become rather complex, and as a consequence, the models can contain many misspecifications, and it is not so simple to cope with this complication.

The use of composite scores based on the multiple indicators occurs more often. In that case, often the Cronbach’s $\alpha$ is computed as an estimate of the reliability of the composite score. These reliability estimates are normally far from 1 and vary in size across the different concepts. Therefore, correction for measurement errors is necessary. These coefficients could be used as quality estimates for the composite scores, and one could apply the procedure mentioned above. However, this is also only seldom done, and if that is not done, one cannot rely on the results of such analyses. Saris and Gallhofer (2007, 2014) indicated that the quality of composite scores can also be evaluated on the basis of information about the quality of single questions, based on SQP, and the correlations between these indicators. This procedure will lead in general to different results with respect to the estimates of the quality of the composite scores because the Cronbach’s $\alpha$ makes more assumptions than this approach.

Thus, we think that the procedure suggested in this paper is simpler and makes fewer assumptions. Therefore, we hope that this procedure will be used more frequently in the future to correct for measurement errors in survey research.

7 Limitations

The new procedure using predictions of the quality of questions is just as good as these predictions are. They depend on the database of questions on which the prediction algorithm has been developed, precision of the estimation of the qualities and the precision of the prediction algorithm. Thus, the corrected results may not be perfect, but it is always better to correct for measurement errors than not to do so, because we know that the quality of the question is never perfect and can lead to serious errors in the estimation. An alternative will be to use SEM, but this should in fact lead to exactly the same estimates (Saris and Gallhofer 2007). Besides, as we can see in Table 1, people are not using it: It seems to be too complicated.

An important limitation of this approach is that MTMM experiments are difficult to formulate for background variables questions and other factual questions asking for dichotomous responses like yes/no or done/not done, etc. However, for these questions,
one can rely on the information about the quality that has been provided in the work of Alwin (2007) based on panel data with the same questions.

Furthermore, we have to say that there will always be new questions that are quite deviant from the questions that are now in the database. For example, so far the new Web survey questions have not yet been entered in the system. So occasionally SQP cannot be used. This means that SQP needs an update regularly to accommodate new question types.

It can also happen that in correction of correlations for measurement errors one obtains values larger than 1 or smaller than zero. This is of course not an acceptable result. The reason for this problem and the way to cope with this situation can be found in De Castellarnau and Saris (2014).

In the analysis of the data using the two-step procedure for correction for measurement errors, one should realize that the standard errors of the estimated regression coefficients and parameters of causal models will be underestimated. If one would use latent variables models with for each latent variable one observed variable with the loading fixed on $q$ and the errors variance on $1 - q^2$ then the estimates of the causal effects would be identical, but the standard errors would be a bit larger. However, also these standard errors would be a bit underestimated. For more details of this issue, we refer to Oberski and Satorra (2013).

On the basis of the presented results and the limitations mentioned, we draw the conclusion that researchers using survey data have in general the possibility to correct for measurement errors in their data and have to do so in order to make their results and conclusions believable.

### Appendix 1

See Table 7.

| Without correction for measurement error | With correction for measurement error |
|------------------------------------------|--------------------------------------|
| Effects on liberal democracy in the UK   | Effects on liberal democracy in the UK |
| Data $ni = 4$, $no = 378$, $ma = km$     | Data $ni = 4$, $no = 378$, $ma = km$ |
| $km$                                     | $km$                                 |
| $1.0$                                    | $.79                                 |
| $.495, 1.0                               | $.495 , .51                          |
| $.401, .413, 1.0                         | $.401, .413, .62                     |
| $.210, -.053, -.116 1.0                  | $.210, -.053, -.116, .92             |
| Labels                                   | Labels                               |
| Liberal just direct income               | Liberal just direct income           |
| Model $ny = 1$, $nx = 3$                 | Model $ny = 1$, $nx = 3$             |
| Out                                      | Out                                  |

$ni$ number of input variables, $no$ number of observations, $ma$ matrix analyzed, $km$ correlation matrix, $ny$ number of dependent variables, $nx$ number of explanatory variables
Appendix 2: The different questions

Now some questions about people from other countries coming to live in Ireland.

B35 How, using this card, to what extent do you think Ireland should allow people of the same race or ethnic group as most Irish people to come and live here? [CARD 14]

| Allow many to come and live here | Allow some | Allow a few | Allow none | (Don't Know) |
|--------------------------------|-----------|------------|-----------|-------------|
| □1                             | □2        | □3         | □4        | □5          |

B36 How about people of a different race or ethnic group from most Irish people? Still use this card. [CARD 14]

| Allow many to come and live here | Allow some | Allow a few | Allow none | (Don't Know) |
|--------------------------------|-----------|------------|-----------|-------------|
| □1                             | □2        | □3         | □4        | □5          |

B37 How about people from the poorer countries outside Europe? Use the same card. [CARD 14]

| Allow many to come and live here | Allow some | Allow a few | Allow none | (Don't Know) |
|--------------------------------|-----------|------------|-----------|-------------|
| □1                             | □2        | □3         | □4        | □5          |

B38 Would you say it is generally bad or good for Ireland's economy that people come to live here from other countries? Please use this card. [CARD 15]

| Bad for the economy | Good for the economy | (Don't Know) |
|---------------------|-----------------------|-------------|
| □0                  | □1                    | □2          | □3        | □4         | □5 | □6 | □7 | □8 | □9 | □10 | □11 |

B39 And, using this card, would you say that Ireland's cultural life is generally undermined or enriched by people coming to live here from other countries? [CARD 16]

| Cultural life undermined | Cultural life enriched | (Don't Know) |
|--------------------------|------------------------|-------------|
| □0                      | □1                     | □2          | □3        | □4         | □5 | □6 | □7 | □8 | □9 | □10 | □11 |

B40 Is Ireland made a worse or a better place to live by people coming to live here from other countries? Please use this card. [CARD 17]

| Worse place to live | Better place to live | (Don't Know) |
|---------------------|----------------------|-------------|
| □0                  | □1                   | □2          | □3        | □4         | □5 | □6 | □7 | □8 | □9 | □10 | □11 |

References

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

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement. The influence of question and respondent attributes. *Sociological Methods and Research, 20*, 139–181.

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

Belson, W. (1981). *The design and understanding of survey questions*. London: Gower.

Biemer, P. R. (2011). *Latent class analysis of survey errors*. Hoboken: Wiley.

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

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

De Castellarnau, A., & Saris, W. E. (2014). A simple way to correct for measurement errors. European Social Survey Education Net (ESS EduNet). http://essedunet.nsd.uib.no/cms/topics/measurement/

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. Hoboken: Wiley.

Goldberger, A. S., & Duncan, O. D. (Eds.). (1973). *Structural equation models in the social sciences*. New York: Seminar Press.
Hagenaars, J. (1988). Latent structure model with direct effects between indicators; local dependency models. *Sociological Methods and Research, 16*, 379–405.

Hambleto, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

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

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Academic Press.

Költringer, R. (1993). *Gültigkeit von Umfragendaten*. Wien: Bohlau.

Költringer, R. (1995). Measurement quality in Austrian personal interview surveys. In W. E. Saris & A. Münnich (Eds.), *The multitrait-multimethod approach to evaluate measurement instruments* (pp. 207–225). Budapest: Eötvös University Press.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.

Lord, F., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesly.

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. The Hague: Mouton.

Molenar, N. I. (1986). *Formuleringseffecten in survey-interviews*. PhD thesis, Amsterdam: Free University.

Oberski, D. T. (2011). Measurement errors in comparative surveys. PhD thesis, University of Tilburg.

Oberski, D. T., Gruner, T., & Saris, W. E. (2011). The prediction procedure the quality of the questions based on the present data base of questions In W. E Saris, D. Oberski, M. Revilla, D. Zavalla, L. Liljefors, I. Gallhofer, & T. Gruner. (Eds.), The development of the Program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working paper 24, chapter 6*.

Saris, W. E, Oberski, D., Revilla, M., Zavalla, D., Liljefors, I., Gallhofer, I., & Gruner, T. (2011). The development of the Program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working paper 24*.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Education Research.

Saris, W. E., & Gallofier, I. N. (2007). *Design, evaluation and analysis of questionnaires for survey research*. Hoboken: Wiley.

Saris, W. E., & Gallofier, I. N. (2014). *Design, evaluation and analysis of questionnaires for survey research* (2nd ed.). Hoboken: Wiley.

Saris, W. E., Satorra, A., & Coenders, G. (2004). A new approach for evaluating quality of measurement instruments. *Sociological Methodology, 3*, 311–347.

Oberski, D. T., & Satorra, A. (2013). Measurement error models with uncertainty about the error variance. *Structural Equation Modeling, 20*, 409–428.

Scherpenzeel, A. C. (1995). *A question of quality. Evaluating survey questions by multitrait-multimethod studies*. Leidschendam: KPN Research.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude survey: Experiments on question form, wording and context*. New York: Academic Press.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey Bass.

Torgerson, W. S. (1958). *Theory and methods of scaling*. London: Wiley.

Tourangeau, R., Rips, J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, MA: Cambridge University Press.

Van Schuur, W. H. (1997). Nonparametric IRT models for dominance and proximity data. In M. Wilson, G. Engelhard Jr, & K. Draney (Eds.), *Objective measurement: Theory into practice* (Vol. 4, pp. 313–331). Greenwich: Alex Publishing Corporation.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33*, 213–239.

Wiley, D. E., & Wiley, I. A. (1970). The estimation of measurement error in panel data. *American Sociological Review, 35*, 112–117.