UNIVERSITAT POMPEU FABRA

RECSM

Research and Expertise Centre for Survey Methodology

RECSM Working Paper Number 31

2013

Universitat Pompeu Fabra - Research and Expertise Centre for Survey Methodology
Edifici ESCI-Born - Despatxos 19.517-19.527
Passeig Pujades, 1 - 08003 Barcelona
Tel.: +34 93 542 11 63 web: http://www.upf.edu/survey/ e-mail: recsm@upf.edu

Correction for measurement errors in survey research: necessary and possible

Willem E. Saris

Melanie Revilla

RECSM/UPF

Abstract

Survey research is the most frequently used data collection method in Sociology, Political Science, Communication, Opinion research and Marketing (Saris and Gallhofer 2007). Nearly everybody agrees that such data contains serious measurement errors. However, only very few researchers try to correct for these errors. If the measurement errors in the different variables are not the same, the comparison of the sizes of effects of 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 measurements errors across variables, methods and countries (Alwin 2007, Saris and Gallhofer 2007). Therefore, correction for measurement errors is essential. This correction can be done in a simple way, but it requires that the sizes of the error variances are known for all observed variables. Within the context of the European Social Survey (ESS), an approach has been developed to solve this problem. In each ESS round 4-6 experiments are done in many different European countries and languages. For the questions involved in these experiments the quality has been estimated and the characteristics of the questions have been coded. This allows the study of the relationship between these characteristics and the quality of the questions. Because this relationship is rather good one can also use it to predict the quality of new questions. This makes the necessary correction for measurement error in survey research possible and easy.

Keywords

Correction for measurement errors, quality, MTMM, SPQ 2.0, European Social Survey

Number of words, excluding figures, tables, references, and appendices: 5105

Introduction

Most studies that require information of individual persons about values, attitudes, opinions, evaluations, feelings, preferences, expectations, status, occupation, education income and behavior rely on interviews or questionnaires. Therefore it is not surprising that it has been found that 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 in depth by many researchers; to mention some important contributions: Belson (1981), Schuman and Presser (1981), Sudman and Bradburn (1983), , Andrews (1984), Alwin and Krosnick (1991), Molenaar (1986), Költringer (1993), Scherpenzeel(1995), Tourangeau et al (2000), Dilmann (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 value of the variables we want to measure is.

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 error and the correction for these errors. Table 1 summarizes these results.

* Table 1 about here*

This table shows 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 have 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 size of the measurement errors and their consequences are relatively small so that 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 size of the measurement errors, or the complement of that, the quality of survey measures, is not available and so correction is not possible.

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

1. Can measurement errors in survey research be ignored?

[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, 9 mentioned the problem of measurement error but only one makes corrections for measurement errors. We are very grateful for this information. It confirms that this phenomenon not only occurs in the Western World but is more general.

In several studies (Andrews 1984, Alwin and Krosnick 1991, Költringer 1995, Scherpenzeel 1995, 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 error. 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 considerable influence 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 (f1) and life satisfaction (f2) represented by the coefficient $\rho(f_1, f_2)$. This coefficient cannot be obtained directly by research. One can only estimate the relationship ($\rho(y_1 y_2)$) 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 strength of the relationships between the variables of interest and the observed variables is the square of the quality coefficient (q_i) of the measurement which may be expressed in a number between 0 and 1, where 0 indicates no relationship at all and 1 indicates a perfect relationship.

Figure 1 about here

It can be shown that the following relation exists between the observed correlation ($\rho(y_1 y_2)$) and the relationship between the variables of interest ($\rho\ f_1 f_2$):

$$\rho(y_1,y_2) = \rho(f_1,f_2)\, q_1\, q_2 \qquad \text{[equation 1]}$$

It will be clear that the two correlations are only equal if the quality of both measures is perfect (1.0), i.e., there are no measurement errors. Unfortunately this will never occur. What happens if the quality of the measures is different from 1.0 is presented in Table 2.

Table 2 bout here

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, the observed correlation will also go down but much faster. If the quality of the measures is on the level of .5, suggested as the average quality in survey research, then the quality coefficients (q) are .7 and the observed correlation will be only half of the size of the correlation between the variables of interest. If the coefficient goes down to .6, then the observed correlation is as small as a third of the true value. It is clear that 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 error or measurement quality also makes comparisons of correlations impossible. 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. It will be clear that 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 then that they have a quality coefficient of .6. In Figure 2 we have presented this situation.

Figure 2 about here

The observed correlations between LS and Age will be .24 (=.4x.99x.6) and between LS and job satisfaction it will be .22 (=.6x.6x.6). Apart from the fact that, due to errors, these correlations are much lower-than they really are, the researcher may also draw the wrong conclusion that age is a bit more correlated than job satisfaction with life satisfaction. This conclusion is wrong because the correlations between the variables of interest are very different, namely, for age .4 and for job satisfaction .6.

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. This point is illustrated in Figure 3.

Figure 3 about here

This Figure shows a situation where the correlations between the observed variables are very different (.65 in country A versus .40 in country B) while the correlations between the latent variables of interest are identical (.80). The reason for these differences is the difference in quality of the measures in the two countries (.90 in country A versus .70 in country B).

The conclusion should be that one has to determine the size of the measurement errors (or the quality) of all variables in the study in order to be able to get unbiased estimates of the relationships between these variables.

Without correction for measurement errors, one runs the risk of very wrong conclusions with respect to correlations between variables and differences in correlations across countries.

2. Is it difficult to correct for measurement errors?

The next issue we would like to discuss is whether it is difficult to correct for measurement errors. In principle the solution to this problem was presented in 1971 when the Structural equation modeling was introduced (Duncan and Goldberger) and where Jöreskog introduced the LISREL program for estimation of such models. A simple example of such a model is presented in Figure 4. Here the researcher have made a model for explanation of Environmental friendly behavior using two endogenous variables "Environmental values" and "Influence" and two exogenous variables : "Reception 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.

Figure 4 about here

Essential in this approach to correction for measurement errors is that one needs for each latent variable at least two observable indicators. The problem of this approach is that one needs at least twice as many observed variables as one has latent variables. This increases the costs of research considerably. It also increases the length of the survey, the burden for the respondents and the complexity of the models and the estimation and testing of these models. These two reasons have been enough to reduce the use of this approach considerably through time even though the procedure is in principle correct.

However, there is also simpler way to correct for measurement errors. This approach is based on equation 1 we have given before. Because if this formula can be formulated then it also true that:

$$\rho(f_1 f_2) = \rho(y_1 y_2)/\, q_1 q_2 \qquad \text{[equation 2]}$$

This result was already known in psychology for a long time (see for example Lord and Novick 1968). So correction for measurement error in the observed correlation is very simple if we know the quality of the observed variables. This result holds for single questions as well as composite scores.

Let us illustrate this procedure by 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. Using Mokken scaling the scores on two latent variables were obtained:

- one based on opinions about liberal rights called “liberal democracy” and
- one based on opinions about electoral requirements called “electoral democracy”

The quality (reliability) of these two scales turned out to be .79 for liberal democracy (q1) and .77 for electoral democracy (q2) while the observed correlations ($\rho(y_1y_2)$) between the two scales was .638. To correct for measurement errors in this correlation we use equation 2 and we get:

$\rho(f_1f_2) = .638/\sqrt{(.79x.77)} = .82$

We see that in this case the correlation by correction for measurement error increases 20%. In the example discussed above it is expected that the scale of 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. 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.

We will now show how correction for measurement error in regression can be done using first the program Lisrel and afterwards the program Stata.

In Table 3 the procedure using the program LISREL has been presented. Table 4 presents the results of this approach. Table 3 shows that the only difference between the input without correction for measurement error and with correction for measurement error is that on the diagonal of the correlation matrix are 1 in the former case and quality estimated in the latter case.

Tables 3 and 4 about here

In table 4, we see that the program computes the correlation matrices correcting for measurement errors using the formula mentioned in equation 2. This leads to considerable differences in the correlations between the variables. As a consequence the estimated effects of the different variables also change considerably by this correction for measurement errors.

While without correction for measurement error, all three variables have significant effects on the opinion about liberal democracy, after correction for measurement errors in these variables 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.

This approach can be used for more complex model correcting for measurement errors in all variables and providing standardized and unstandardized coefficients.

Let us now illustrate how correction can be done using Stata. We will show that in that case the correction is also very simple but the possibilities are at this moment more limited. One can apply it only on regression, not on causal models in general. Only correction for measurement error in the independent variables is possible. Therefore, one can only get the unstandardized coefficients. Nevertheless, it is interesting to illustrate how simple the procedure is and how large the differences are. In table 5, the analysis without correction for measurement errors is presented and in Table 6 the analysis with correction for measurement errors.

Tables 5 and 6 about here

In Table 5 the command “regress” is used whereas in Table 6 the command “eivreg” is used. In the latter case, the qualities of the indicators have to be indicated. The results are again very striking. While without correction for measurement error the regressions coefficients (unstandardized) of “socjustice” (former “Just”) and “Direct” are approximately equal the difference is after correcting for measurement error nearly a factor 10 and the effect of “Direct” is not significant anymore. This illustrates how big the effects of correction for measurement error are on the results of a simple regression analysis. Moreover, what is here at least as important is that we see that in both programs, the correction for measurement error is very simple. Given the huge effect of the results of the analysis one would say that one cannot report analyses without correction for measurement error.

3. Are estimates of the quality of survey measures missing?

There are a lot of different procedures to estimate the quality of questions and 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 was the Quasi simplex model (Heise 1969, Wiley and Wiley 1970) used by Alwin and Krosnick (1971) and (Alwin 2007). The Mulitrait-Multimethod or 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, Harman1976 and latent class analysis Hagenaars (1988, Vermunt (2003), Biemer (20011). Besides that scaling methods have been developed like Thurstone scale, Likert scale etc (Torgerson 1958) , Gutmann scale and Mokken scale (1969), unfolding scale (Van Schuur 1997) Rasch scale Rasch (1960) and Item Response theory (Hambleton et al. 1991). For the advantages and disadvantages of these different procedures we refer to this literature.

All these procedures require at least 2 questions for each concept. That means that the number of questions is at least twice the number of concepts one likes to take into account in the analysis. 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 or concepts specific for the formulations of these questions used in the specific questionnaire and context. Generalization is not easily possible.

This means that a lot of research has to be done before the final data collection in order to correct for measurement errors in all variables to be used. This is so much work that it is only seldom done as we have seen in Table 1. So the question is whether there is a procedure that is less time and money consuming to estimate the quality of survey questions and of composite scores for concepts with multiple indicators.

A new approach

From the very start of the European Social Survey, the author of this paper, as a member of the Central Coordinating Team (CST), has emphasized that their measures will contain errors and, that without correction for these errors, the results would be questionable and incomparable across countries. Therefore, from the start of 2002, each round of the ESS survey contains 4 to 6 experiments to evaluate the quality of the questions. These experiments were carried out in most countries and all rounds. Consequently, after 5 rounds, more than 600 experiments in more than 20 countries (languages) have been completed, involving approximately 4000 questions.

However, this information is not enough because in the same time in the ESS more than 60.000 questions were asked with respect to values, opinions, attitudes, preferences, feelings etc. So a different approach was required.

Table 7 about here

The new idea was to code the characteristics of the questions, and with this knowledge to develop an algorithm with which to predict the quality of the questions (Saris and Gallhofer 2007). If that prediction would be successful, the same algorithm could be used to predict the quality of any other question as well. This new approach has been worked out with a subsidy of the European Commission for Infrastructure research, and has led to the development of the program SQP 2.0 that contains quality information of the questions involved in the experiments but can also be used to predict the quality of other questions. For the complete report about the development of this tool we refer to Saris et. al. (2012). Here, we mention only the basic steps that were introduced in this process.

Split ballot MTMM design and model

In the normal MTMM experiment the respondent has to provide responses to three different questions (traits) measured with three different methods (Andrews 1984). Because people had to answer three times approximately the same question, one may expect memory effects. In order to cope with the memory effects in the MTMM experiments, it has been suggested by Saris to randomly split the sample into different subgroups and ask the same question only twice in each group. Saris, Satorra and Coenders (2004) showed that this design also allows the estimation of the reliability and validity (complement of the method effect and the quality of each question[2]). During the last years, all experiments of the first three rounds of the ESS have been analyzed using the procedure developed. This means that we obtained estimates of the reliability, validity and quality of all questions involved in the experiments and in all the different countries and languages.

[2] Quality is defined as the product of the reliability and the validity

The coding of the questions

The idea was to use the characteristics and the context of the questions as predictors of their quality. Therefore, we have made a program to code the questions that were involved in the experiments in the ESS. The characteristics used are summarized in table 7.

For details of these characteristics, we refer to Saris et al (2012). People who were speaking the different languages involved in the ESS and were able to understand English coded the questions. This was a very elaborate task but results were rather rewording as we will mention below.

The prediction model

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 provided rather good predictions of the reliability and validity for our data. The R^2 for reliability was .65 and for validity .84. Also, the prediction of the quality was as a consequence rather good.

Survey quality prediction: SQP

Based on this knowledge, this algorithm has been used to develop the computer program SQP 2.0[3] to generate predictions on the quality of questions (Oberski et al 2012).

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

In principle, this approach solves the major problem for the researchers: that one needs the estimates of the quality of all variables in the study in order to be able to correct for measurement error in the analysis.

Derivation of the quality of complex concepts

So far we discussed the estimation of the quality of single questions. Often researchers use concepts based on several indicators. So we need also a solution for the estimation of the quality of composite scores for complex questions. Such a solution indeed exists based on the evaluation of the estimation of the quality of single questions. For any single question we can formulate:

$$y_i = t_i + e_i \qquad \text{[equation 3]}$$

where y_i is the observed variable, e_i the error of measurement and t_i is just the difference between the two. In general, the composite score is defined as a weighted average of the observed variables:

$$CS = \Sigma_i\, w_i y_i = \Sigma_i\, w_i(t_i + e_i) \qquad \text{[equation 4]}$$

[3] SQP is free of change available at sqp.upf.edu

where CS stands for the composite score, w_i is the weight for the i^{th} observed variable.

The quality of a variable can be defined as the ratio of the systematic variance of the variable and the total variance of the variable or 1 – (error variance / total variance). In this case that would be:

$$\text{Quality of CS} = 1 - (\text{var}(e_{cs})/ \text{var}(CS)) \quad \text{[equation 5]}$$

The total variance of the CS can be obtained directly from the computed composite score while the error variance is equal to:

$$\text{Var}(e_{cS}) = \Sigma w_i^2 \text{ var}(e_i) + 2\Sigma w_i w_j \text{ cov}(e_i e_j) \quad \text{over i and } i \neq j \quad \text{[equation 6]}$$

Where $\text{cov}(e_i e_j)$ is equal to the CMV[4] for the variables y_i and y_j while $\text{var}(e_i)$ = the error variance in y_i and can be estimated as:

$$\text{var}(e_i) = (1-q_i^2)\text{var}(y_i) \quad \text{[equation 7]}$$

This derivation shows that one can obtain an estimate of the quality of a composite score from the estimation of the quality of the single questions and the CMV. This suggests that we can get an estimate of the quality of composite scores for complex concepts, if we are able to solve the estimation of the quality of single questions.

Some conclusions and limitations

At the end of this section we can conclude that nowadays a simple procedure is available for estimation of the quality of existing and new questions, even for complex concepts. The researcher does not have to do any extra research and does not have to extend his/her data collection to at least the double number of questions than the

[4]For the calculation of CMV , see note 5 below

number of concepts in the analysis. The only thing that has to be done is the coding of the involved questions. This can even be done before the data collection in order to detect if the quality of the questions is not too low to be used. If that is the case, the program can also provide suggestions for improvements of the questions.

The only limitation of this approach is that MTMM experiments are difficult to formulate for background questions and other factual questions. 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.

An elaborate illustration

As an illustration of the procedure we have discussed here, we have chosen a research issue that has been studied by many researchers recently on the basis of the ESS data. It is the explanation of opinions of people about extra immigration of foreigners in their country. Some of the variables introduced in the ESS for explanation of this opinion and the model proposed are presented in Figure 5.

Figure 5 about here

The questions asked to measure these concepts are presented in Figure 6.

Figure 6 about here

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 B37 is quite different from the other three which have been specified with the same scale type. This can lead to a standard reaction of the respondents which has been called the method effect. 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. The adjusted model is presented in Figure 7.

Figure 7 about here

In this case, equations 1 and 2 have also to be changed in order to take this common method variance into account. The more realistic equations are presented in equations 8 and 9.

$$\rho(y_{1j},y_{2j}) = \rho(f_1,f_2)q_{1j}q_{2j} + cmv \quad \text{[equation 8]}$$

From which follows as before:

$$\rho(f_1,f_2) = [\rho(y_{1j},y_{2j}) - cmv]/\ q_{1j}q_{2j} \quad \text{[equation 9]}$$

The quality of these questions decomposed in reliability, validity and method effect are presented in Table 8.

Tables 8 and 9 about here

In the top part of Table 9, the observed correlations between these variables have been presented below the diagonal. Above the diagonal the CMV[5] 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 presented. 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 error.

[5] It can be shown that the cmv= $r_{1j}m_{ij}m_{2j}r_{2j}$

The results of these two analyses are presented in table 10. The table shows again the considerable differences in the effects whether one corrects for measurement error or not.

Table 10 about here

While the economic threat has a significant effect in both regression equations if one does not correct for measurement errors, these effects are reduced to zero after correction for measurement error, while the effects of Better life and Cultural threat have 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 error in the analysis.

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 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, in general, that:

1. the relationships between variables are underestimated
2. the estimates of effects of different variables can be very biased and lead to wrong conclusions
3. the correlations across countries cannot be compared.

We hope to have shown that correlation for measurement error is absolutely necessary.

The second possible reason for not correcting for measurement error was that the correction procedures are too complicated. We have indicated that this is indeed the case if we look at the procedures developed in the context of Structural equation modeling and multiple indicator modeling. However we have also shown that very simple procedures are available for correction. We showed that any model can be estimated correcting for measurement errors by substituting for the variances of the variables in the correlation or covariance matrix by the quality of the question. In this case one gets automatically estimates of the parameters in the model corrected for measurement errors. The only disadvantage of this procedure is that the standard errors of the estimates are a bit underestimated because we assume that the quality of the measures is exactly known. For more details of this issue we refer to Saris et al. (2012) and Oberski and Satorra (2013). We have also shown that in Stata very simple procedures are available for correction for measurement error although they are less general than in SEM programs. The general conclusion should be that there are simple procedures for correction for measurement error available. So this cannot be a reason not to correct for measurement errors.

The third possible reason was that the size of the measurement errors or the quality of the measures is not available. It was 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 meta-analysis of a lot of measurement quality experiments which makes it possible to predict the quality of any substantive variable 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. We indicated that for objective variables like background variables and factual information the program is not the proper source but for these variables the work of Alwin (2007) provides the necessary information.

On the basis of these results we draw the conclusion that researchers that use survey data have the possibility to correct for measurement error in their data and have to do so in order to make their results and conclusion believable.

REFERENCES

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

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

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

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

Goldberger A.S and O.D.Duncan ,Eds, (1973) Structural equation models in the social sciences. NewYork , SeminarPress.

Hagenaars J. (1988) Latent structure model with direct effects between indicators; local dependency models. *Sociological Methods and Research*. 379-405

Harman (1976) *Modern factor analysis*. Chicago. University of Chicago Press

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

Költringer R. (1983) *Gültigkeit von Umfragendaten*. Wien, Bohlau

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

Lord F., and M. R. Novick (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*. New York: Walter Gruyter Mouton.

Molenaar. N. I. (1986). *Formuleringseffecten in Survey-Interviews*. *PhD thesis*, Amsterdam: Free University.

Muthen B. (1984) A general structural equation model with dichotomous, ordered and continuous latent variables indicators. *Psychometrika*, 49, 115-132.

Oberski D.T.and A. Satorra (2013). "Measurement error models with uncertainty about the error variance" *Structural Equation Modeling (forthcoming)*

Rasch G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Education Research.

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

Saris W.E. and I.N. Gallhofer (2007) *Design, evaluation and analysis of questionnaires for Survey research*, Hoboken, Wiley.

Saris W.E, D.Oberski, M.Revilla, D.Zavalla, L.Lilleoja, I.Gallhofer and T.Grüner (2011)

The development of the Program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working paper 24*

Scherpenzeel A. C. (1995). *A Question of Quality. Evaluating Survey Questions by Multitrait-Multimethod Studies*. KPN Research: Leidschendam.

Schuman H., and S. Presser (1981). *Questions and Answers in Attitude Survey: Experiments on Question Form, Wording and Context*. New York: Academic Press.

Sudman and Bradburn (1983) *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., J.Rips, K.Rasinski, (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: Wilson, M., Engelhard jr. G., & Draney, K. (eds.), *Objective Measurement: Theory into Practice*, Volume 4. Greenwich (Cn)/London: Ablex Publishing Corporation, 313-331.

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

Wiley D. E., and I. A. Wiley 1970. The estimation of measurement error in panel data. *American Sociological Review*, 35, 112-117.

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

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

Note: ESR=The European Sociological Review, EJPR= European Journal of Political Research, POQ= Public Opinion Quarterly, APSR=The American Journal of Political Science, JM=Journal of Marketing

Table 2: The effect of the measurement quality on the observed correlation given that the correlation between the variables of interest is .9 i.e. $\rho(f_1 f_2)$=.9

Quality coefficient	Quality coefficient	Observed correlation
q_1	q_2	$\rho(y_1, y_2)$
1.0	1.0	**.90**
.9	.9	**.73**
.8	.8	**.58**
.7	.7	**.45**
.6	.6	**.33**

Table 3: Procedure to correct for measurement error using LISREL

Without correction for measurement error	With correction for measurement error
Effects on liberal democracy in the UK	Effects on liberal democracy in the UK
Da ni=4 no=378 ma=km	Da 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
Here 1 on the diagonal	**Here quality on the diagonal**

Table 4: The correlations and regression from LISREL

Without correction for measurement errors

Correlations

	liberal	just	direct	income
liberal	1.00			
just	0.50	1.00		
direct	0.40	0.41	1.00	
income	0.21	-0.05	-0.12	1.00

Regression (36% explained)

	just	direct	income
liberal	0.40	0.27	0.26
s.e.	(0.05)	(0.05)	(0.04)
t-value	8.77	5.84	6.29

With correction for measurement errors

Correlations

	liberal	just	direct	income
liberal	1.00			
just	0.78	1.00		
direct	0.57	0.73	1.00	
income	0.25	-0.08	-0.15	1.00

Regression (70% explained)

	just	direct	income
liberal	0.76	0.07	0.32
s.e.	(0.04)	(0.04)	(0.03)
t-value	18.22	1.59	11.06

Table 5: Regression without correction in STATA

regress liberal socjustice direct income if cntry==1

Source	SS	df	MS
Model	420.412508	3	140.137503
Residual	801.787492	316	2.53730219
Total	1222.2	319	3.83134796

Number of obs	= 320
F(3, 316)	= 55.23
Prob > F	= 0.0000
R-squared	= 0.3440
Adj R-squared	= 0.3378
Root MSE	= 1.5929

liberal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
socjustice	.8973032	.125228	7.17	0.000	.6509172	1.143689
direct	.7825098	.145502	5.38	0.000	.4962346	1.068785
income	.1694866	.0286066	5.92	0.000	.1132031	.22577
_cons	.6686871	.2038819	3.28	0.001	.2675495	1.069825

Table 6: The procedure for correction in STATA

eivreg liberal socjustice direct income , r(socjustice .51 direct .62 income .92), if cntry==1

variable	assumed reliability
socjus~e	0.5100
direct	0.6200
income	0.9200
*	1.0000

Errors-in-variables regression

Number of obs	= 320
F(3, 316)	= 74.96
Prob > F	= 0.0000
R-squared	= 0.5166
Root MSE	= 1.36732

liberal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
socjustice	2.1942	.4621737	4.75	0.000	1.284873	3.103526
direct	.2091936	.4696542	0.45	0.656	-.7148508	1.133238
income	.1870243	.0273677	6.83	0.000	.1331782	.2408703
_cons	-.3465327	.2769938	-1.25	0.212	-.8915179	.1984525

Table 7: The characteristics of the questions that are coded to predict the quality

Group		Specific characteristic
Group 1	The trait	Domain Concept
Group 2	Associated to the trait	social desirability centrality of the topic time specification
Group 3	Formulation of the request for an answer	trait requested indirectly, direct or no request and presence of stimulus (battery) WH word and what type of WH word Type of the request (interrogative, Imperative question-instruction, declarative or none (batteries). Gradation Balance of request or not Encouragement to answer Emphasis on subjective opinion Information about the opinion of other people Absolute or a comparative judgment
Group 4	Characteristics of the response scale	Categories; yes/no answer scale; frequencies; magnitude estimation; line production and, more steps procedures. Amount or the number of categories full or partial labels labels with long or short text Order of labels Correspondence between labels and numbers theoretical range of scales (bipolar or unipolar) Range of scales used Fixed reference points Don't know option
Group 5	Instructions	Respondent instructions Interviewer instructions
Group 6	Additional information about the topic	Additional definitions, information or motivation
Group 7	Introduction	Introduction and if request is in the introduction
Group 8	Linguistic complexity	Number of sentences Number of subordinated clauses Number of words Number of nouns Number of abstract nouns Number of syllables
Group 9	Method of data collection	
Group 10	Language of the survey	
Group 11	Characteristics of the show cards	Categories in horizontal or vertical layout Text is clearly connected to categories or if there is overlap Numbers or letters shown before answer categories Numbers in boxes Start of the response sentence shown on the show card Question on the show card Picture provided.

Table 8: The predicted values of the quality indicators in Ireland

Variable	**Method**	**r^2**	**v^2**	**m^2**	**q^2**
Allow	SQP2.0	.826	.906	.094	.747
Economy	SQP2.0	.770	.780	.220	.601
Culture	SQP2.0	.761	.705	.295	.537
Better	SQP2.0	.748	.725	.275	.543

Table 9: Correction for errors and CMV

Observed correlations with CMV above the diagonal				
Allow	1.0	0.0	0.0	0.0
Better Life	-.470	1.0	**.186**	**.215**
Economic Threat	-.423	.662	1.0	**.195**
Cultural Threat	-.447	.718	.704	1.0
Covariance matrix used in estimation of the model				
Allow	.747			
Better Life	-.470	.543		
Economic Threat	-.423	**.476**	.601	
Cultural Threat	-.447	**.503**	**.509**	1.0

Table 10: Estimates of the parameters with and without correction

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

Figure 1: a very simple model

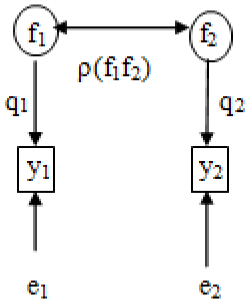

Figure 2: A model of the effects of Job satisfaction (JS) and Age on Life satisfaction (LS) taking into account measurement error

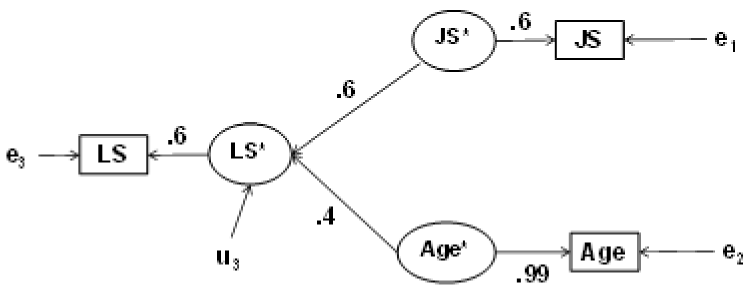

Figure 3: Consequences for cross-cultural comparisons

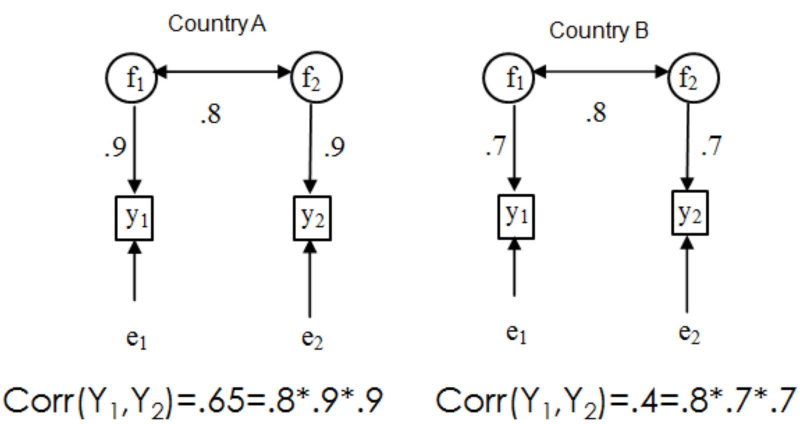

Figure 4: The standard SEM approach

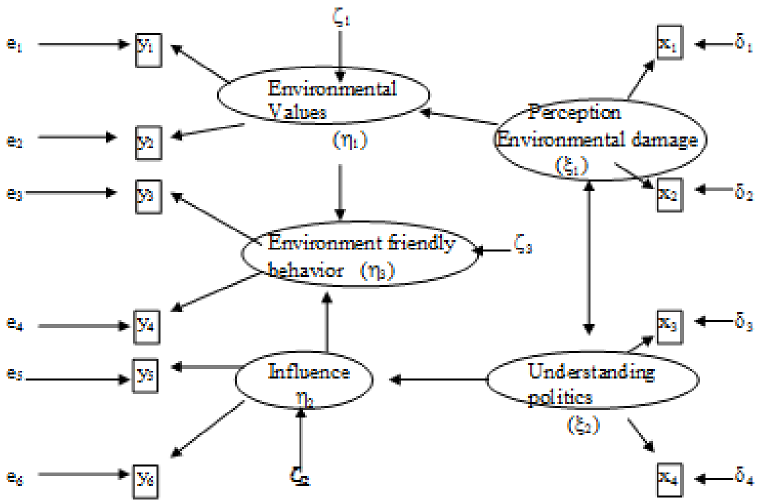

Figure 5: A simple model for the explanation of the opinion about immigration of people from outside Europe

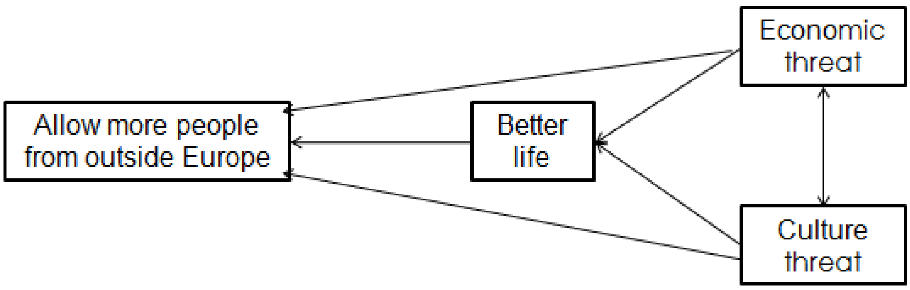

Figure 6: The different questions

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

B35 Now, using this card, to what extent do you think Ireland should allow people of the <u>same race or ethnic group</u> 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	☐8

B36 How about people of a <u>different</u> 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	☐8

B37 How about people from the <u>poorer countries outside Europe</u>? 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	☐8

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 ☐88

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 ☐88

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 ☐88

Figure 7: Correction for errors taking into account CMV