Survey Research Methods (2017)
Vol. 11, No. 4, pp. 369-382
doi:10.18148/srm/2017.v11i4.7170

© *European Survey Research Association*
ISSN 1864-3361
http://www.surveymethods.org

The relative size of measurement error and attrition error in a panel survey. Comparing them with a new Multi-Trait Multi-Method model.

Peter Lugtig
Department of methodology and statistics
Utrecht University
The Netherlands

This paper proposes a method to simultaneously estimate both measurement and nonresponse errors for attitudinal and behavioural questions in a longitudinal survey. The method uses a Multi-Trait Multi-Method (MTMM) approach, which is commonly used to estimate the reliability and validity of survey questions. The classic MTMM model is in this paper extended to include the effects of measurement bias and longitudinal nonresponse that occurs in longitudinal surveys. Measurement and nonresponse errors are expressed on a common metric in this model, so that their relative sizes can be assessed over the course of a panel study. Using an example about political trust from the Dutch LISS panel, we show that measurement problems lead to both small errors and small biases, that dropout in the panel study does not lead to errors or bias, and that therefore, measurement is a more important source of both error and bias than nonresponse.

Keywords: Trade-off between nonresponse and measurement; MTMM; panel study; total survey error; nonresponse

1 Introduction

Two of the most studied sources of error in surveys are measurement error and nonresponse error. In absence of validation data, researchers rely on population or sampling frame data to assess nonresponse error (Groves, 2005), and statistical models to estimate measurement error (Alwin, 2014). Because of the fact that different methods are used to estimate nonresponse and measurement error, the metrics of these survey errors are different. We cannot compare the relative sizes of both types of error, and therefore cannot say whether measurement error or nonresponse error contributes most to total survey error. As a consequence, we cannot study interactions between measurement and attrition errors either, limiting the progress survey methodologists can make in designing surveys that minimize total survey error.

Nonresponse and measurement error in surveys are often believed to interact. 'Difficult' to recruit respondents have for example been found to report with more measurement error than 'easy' respondents (Cannell & Fowler, 1963; Fricker & Tourangeau, 2010; Kaminska, McCutcheon, & Billiet, 2010; Olson, 2013). The same may hold true for the relation between measurement error and attrition. Those reporting with more measurement error, may be at a higher risk of dropout later in the panel survey.

The relation between measurement error and attrition may be attributable to a common cause: respondents may not be motivated in general or find it difficult to complete the surveys, leading to either measurement error, attrition from the survey, or both.

This paper concentrates on investigating such a relation between measurement and longitudinal nonresponse. The goals are twofold: First, to illustrate a new method to quantify the effects of measurement and longitudinal nonresponse error on the same metric in order to meaningfully compare the two and their relation. Second, using the LISS panel study as an example, to assess the relative size of measurement and attrition error in a longitudinal survey, and study whether the two error sources interact.

1.1 Background

Longitudinal nonresponse (from here on called attrition) and measurement error are believed to be two of the largest sources of error and bias in surveys (Lynn, 2009). Measurement error occurs when respondents by accident or on purpose give an answer to a question that is not their 'true' answer. Attrition occurs when respondents do no longer participate in one or multiple measurements that are conducted in a longitudinal study. Attrition can be temporary or permanent, and can bias survey estimates when people who drop out are different from continuing respondents.

Contact information: Peter Lugtig, Department of Methodology and Statistics, Utrecht University, Padualaan 14, 3584 CH, Utrecht, The Netherlands (E-mail: p.lugtig@uu.nl)

Great efforts and costs are spent to limit attrition and measurement error in panel surveys. Measurement error can be reduced by using validated survey questions and pretesting those before fielding the survey. Attrition is usually limited by sending advance letters, using incentives, keeping in touch with panel members or following up wave-nonrespondents (Couper & Ofstedahl, 2009; Watson & Wooden, 2009). Further design features that affect the size and trade-off between attrition and measurement error are for example the choice of survey mode, the use of interviewers, and the question topics. We know theoretically that each of the choices we make in designing a survey affect both attrition and measurement error to a different extent. For example, the choice of a self-administered survey mode (e.g. mail or Internet), is generally assumed to limit measurement error in comparison to interviewer administered surveys, but lead to lower response rates (Dillman et al., 2009). Design choices also affect survey errors other than attrition and measurement (mainly sampling, coverage and adjustment error), but this papers focuses on attrition and measurement error in the context of a longitudinal survey only.

Remarkably, we know little about how these design choices affect total survey error, making it hard to allocate resources to limit attrition or measurement error (Lynn & Lugtig, 2017). Several validation studies in recent years have tried to study the relative contributions of nonresponse and measurement error for different questions and survey modes (Kreuter, Müller, & Trappmann, 2010, 2013). Felderer, Kirchner, and Kreuter (2013) for example use administrative records to study errors in social demographic variables, and benefit receipt in Germany in a randomized Web/Telephone study. They find that nonresponse error is larger in the Web-than in the telephone survey, and that in comparison, nonresponse error contributes more to total survey error than measurement error, even for a socially desirable variable like benefit receipt. Sakshaug, Yan, and Tourangeau (2010) draw similar conclusions in a validation study among college students, and note that nonresponse and measurement errors usually reinforce each other. The fact that nonsampling errors generally appear to go in the same direction lead to the conclusion that total survey errors can be large, or very large.

Apart from the few studies just discussed that used factual validation data, we know little about whether it is measurement error or attrition error that contributes most to total survey error. This in turn makes it impossible to efficiently allocate resources towards limiting either error source, or in fact make a choice between several survey design features.

The aim of this paper is to show how attrition and measurement error for any variable can be expressed on a common metric, when the two are studied in the context of a panel survey. The natural response scale of a variable will serve as the common metric, meaning that bias and errors will be expressed as the deviances in means and (co)variances between the observed data and estimates from a statistical true-score model (the Multi-Trait Multi-Method model).

The remainder of this paper is structured as follows. First, the possible effects of attrition and measurement error on key survey estimates are described. Second, the MTMM design is introduced as a method to estimate measurement error, in the form of the reliability and validity of survey questions. In order to assess the effects of attrition as well, the classic MTMM model will be extended to include means as a form of measurement bias, and include the effects of attrition by estimating the MTMM model with means, in a multi-group Structural Equation Model. Next, the results for this model are discussed using data from the Dutch LISS panel. We show how the model can be used to estimate and compare the sizes of attrition and measurement error for 9 questions that ask respondents about their political trust. Both attrition errors and measurement errors in the LISS panel turn out to be small. Further, attrition and measurement error do not interact – implying that there is no relation between the amount of measurement error respondents report with, and their propensity to drop out of the survey at a later stage. We conclude with a discussion of how our approach may be used in panel surveys to inform survey designs, as well as the limitations and several possible extensions of the approach.

1.2 Assessment of survey errors

Survey errors have a variable and systematic component. Variable errors affect the reliability of a survey estimate (random error), while systematic errors affect the validity (bias) (Biemer, 2010). Each component of total survey error has a variable and systematic error component, which all result in increased variance or bias in key survey estimates. Within the total survey error framework, Biemer (2010) has recommended to study Mean Squared Error (MSE) as a means to study the size of the combined errors and bias.

The sizes of attrition and measurement errors and biases are much more difficult to estimate than for example sampling errors. The nature of attrition will vary with every project, depending on the survey topic, population, and design. Whether attrition mainly leads to variable or systematic error can only be assessed when complete information for either the sample frame or the population as a whole is available. When population or sample frame level data are available for all respondents, they can also be used to evaluate measurement error. Typically, researchers have such validation data only for factual or behavioural variables, and not for attitudinal variables, which are often important outcome measures of surveys. In absence of validation data, researchers therefore have to rely on statistical modelling and use a different metric than MSE to assess the size of survey errors.

2 MTMM Models

The Multi-Trait Multi-Method approach (MTMM) was proposed by Campbell and Fiske (1959), and later applied to study the quality of survey questions by for example Saris and Andrews (1991). The idea of the model is that the same concept of interest (in this paper political trust) is measured with 9 questions that differ in their content (trait), and question formats (methods). For traits, different but related concepts are measured – for example different aspects of social trust or media consumption (Saris & Gallhofer, 2007). As methods, various types of response scales, the number of response options, the presence of labels, the availability of a don't know answer, and questionnaire introductions have been used in the past. From these studies we have learnt a lot about what specific question formats work best for which types of traits (Saris & Gallhofer, 2007; Scherpenzeel & Saris, 1997).

Although there are many different ways to analyse MTMM data, Figure 1 shows the classic MTMM model, which in the literature is often referred to as the correlated-trait orthogonal methods model (Widaman, 1985).

The power of the MTMM lies in the fact that it can both estimate the convergent and divergent validity (Alwin, 2014). The 9 different questions combining 3 traits and 3 methods of an MTMM-model are usually highly correlated. Those questions that measure the same trait, or are using the same method, are however more strongly correlated than questions which use a different trait and method. This fact is used to decompose the total variance into different components using the model shown in Figure 1, or as expressed in equation 1. Note that in the context of a panel study, the MTMM model estimates measurement error at one cross-section. Often, one is also interested in measurement errors of change estimates, but this is beyond the scope of this paper.

The basis for the MTMM model is the notion that every observed variable Y_{jk} is the result of the Latent Trait (T_j) and random error (e_{jk}) for a certain trait (j) and method (k) similar to what is assumed in true-score theory (Lord, Novick, & Birnbaum, 1968) and what is called the classical MTMM-model by Saris and Andrews (1991). The degree to which the observed score Y_{jk} is determined by the Latent Trait is determined by the indicator validity coefficient (λT_{jk}) of the score. In the MTMM, the observed scores Y do not only depend on the Latent Trait however, but also on the Method (M_k) that is being used to ask a question. The degree to which the method determines the observed score is expressed in an invalidity coefficient or method-effect coefficient (λM_{jk}), so that

$$Y_{jk} = \lambda T_{jk} \cdot T_j + \lambda M_{jk} \cdot M_k + e_{jk} \quad , \tag{1}$$

for each j, k.

Two assumptions are necessary to empirically test the MTMM. First, the random errors (e_{jk}) are uncorrelated with each other. Second, the method (M_k) and trait factors (T_j) are uncorrelated. Method factors are usually uncorrelated among themselves, and the variances of all factors are set to 1, although these last two restrictions are not always implemented (Saris & Andrews, 1991).

2.1 MTMM model with means

In the classic MTMM model as depicted in Figure 1, the effects of the MTMM design on the (co)variances are modelled. Means are usually ignored. The different question formats used in the MTMM design can however also lead to substantial higher or lower means in the data.

The problem with the model outlined in equation 1 is, that it is impossible to estimate both the observed means (Y_{jk}) and latent means (T_j, M_k) simultaneously. Coenders and Saris (2000) have first outlined how method means (M_k) can be estimated in the model, while Pohl and Steyer (2010) formalized different ways to do this. The MTMM model with means used in this paper is what Pohl and Steyer (2010) call the "Method effect model with a reference model" (MEref; see also Pohl, Steyer, & Kraus, 2008). Here, the intercepts (Y_{jk}) of all observed variables, as well as one of the Latent method factor means (M_1) are set to zero. Then, all observed means can be decomposed into 3 Latent trait means (T_j), which are similar to a latent mean in a normal CFA model, and 2 relative Latent method means ($M_{2,3}$) which indicate the difference in the Latent Means as compared to the reference method. The estimated latent mean (α) and variance (ψ)of the reference method (M_1) are constrained to 0, and so are the loadings. For this reason, the reference method (M_1) is absent from Figure 2, and all parameters shown for M_2 and M_3 indicate relative differences from M_1. In other words, the latent method means indicate how much, on average, the means of the survey questions would shift when a particular survey method is used instead of the reference method. Further constraints to identify the model are imposed on λT_{jk} and λM_{jk}. These constraints however do not affect the ability to estimate random and systematic errors in the covariance structure.

The MTMM model as specified above in figure 2 accommodates both random and systematic measurement error in two specific ways. It estimates how variances and covariances are attenuated because of the reliability and validity of the questions, and how means are biased because of a method effect. In short, the MTMM model with means can accommodate all possible effects of measurement errors. Because the mean and variance of the reference method are restricted, we have chosen the Method Factors in the model above not to be correlated, which seems warranted when correlations between method factors are not assumed to be high (Conway, Lievens, Scullen, & Lance, 2004; Revilla & Saris, 2013). In theory this can be done, but in practice we have often encountered convergence problems as reported in Pohl and Steyer

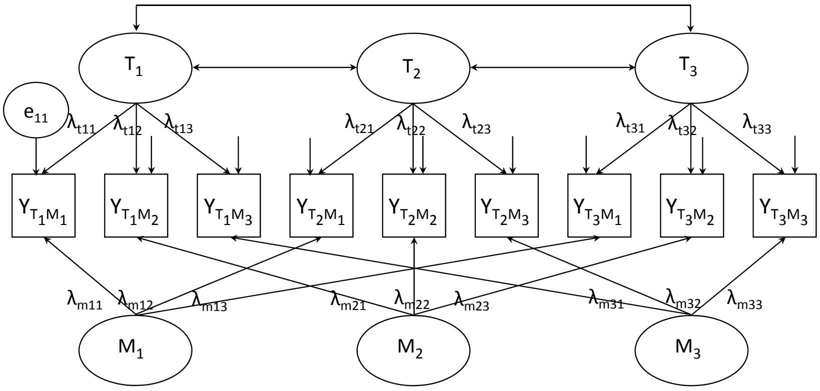

Figure 1. Traditional MTMM model with three methods and three traits. For simplicity, only the residual for e_{11} is shown

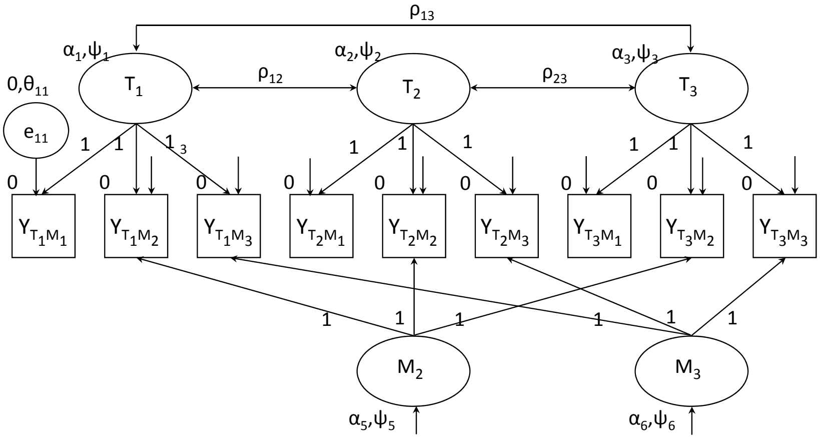

Figure 2. MTMM model with means and parameter constraints (MEref model). Constraints for residual measurement errors only shown for e11.

(2010) as so we refrained from using models with correlated Methods factors

2.2 Extension to attrition error (longitudinal nonresponse)

The MTMM model with means as shown in figure 2 is a cross-sectional model. It estimates the effects on validity and reliability for 3 traits and 3 methods, and relative bias in means for 2 methods. In order to assess the effects of longitudinal nonresponse as well, the MTMM model needs one further extension to a longitudinal component. This new version of MTMM model will enable us to express the effects of measurement and nonresponse error on the same metric.

Respondents who drop out from the survey can be systematically different from respondents who stay in, leading to attrition error or bias. The form of the bias can again differ; attrition may lead to differences in means, or (co)variances. An easy way to extend the MTMM model is to separate the respondents and nonrespondents in later waves of the panel, and analyse a series of MTMM models with means for the two groups separately. Hox, de Leeuw, and Chang (2012) used a similar approach to study the measurement error for 'eager' and 'reluctant' respondents during the recruitment phase of a survey.

Depending on how many waves of data are available, the process of attrition can be technically modelled in a parametric (e.g. timeseries or multilevel), or nonparametric way (every wave as a separate model). Because we have no theoretical idea of how the process of attrition affects the parameters in the MTMM model, we believe it is easiest to model attrition as a series of separate models. The MTMM model is split for 1) respondents and 2) nonrespondents in a particular wave ($t + p$), and subsequently estimated using multi-group modeling in each wave. The differences between the MTMM model with complete data (wave t), and restricted to later respondents (wave $t + p$) then informs us about the effects of attrition. This model includes two extra assumptions on how measurement error and time interact:

1. Measurement errors are assumed to be constant over time for all respondents. The MTMM model is administered only once, and so we assume that respondents would have responded in a similar way had the MTMM questions been repeated in later waves. There is evidence that MTMM estimates are indeed stable when administered repeatedly, so we are confident about using this assumption (Koch, Schultze, Eid, & Geiser, 2014).

2. The systematic bias in means in the reference method is assumed to be stable over time. Because the difference in the method means is a relative difference, changes in the relative difference over time can either be caused by a change in the reference method, or the comparison method. Also, a systematic shift across all three methods over time cannot be detected. This is arguably a stronger assumption, and leads to the conclusion that the estimates of the change of the relative method effect over time have to be interpreted with some caution.

It is important to keep in mind that only attrition which occurs after the MTMM has been administered can be modelled. And while our method can be used to compare the size of attrition and measurement error, it can only do so for attrition after wave 1, thereby excluding nonresponse in the panel recruitment stage.

3 Example

To show how the MTMM model with means can be used to assess the size and relation between attrition and measurement error in a panel survey, we use data from the Longitudinal Internet Study for the Social Sciences (Longitudinal Internet Study for the Social sciences (LISS), 2008), run by CentERdata, at Tilburg University, the Netherlands. This panel study started at the end of 2007 with a simple random sample of almost 17,000 individuals taken from community registers. Potential respondents were contacted by letter, telephone or in-person visit, and after an initial interview ("recruitment stage") were asked to become a member of the online panel (which they start with a "profile interview"). Although the LISS panel is Internet-based, it was not necessary to own a personal computer with an Internet connection to participate in the panel, as CentERdata provided the equipment if required. Using the response metrics of Callegaro and Disogra (2008), the recruitment rate (or RECR, similar to AAPOR RR3, defined as the number of people that agree to join the panel, relative to all people invited) for the LISS panel is 63 per cent. The profile rate (or PROR; defined as the number of people that have completed the profile interview, relative to all people invited) is 48 per cent. Retention is about 90% a year (Binswanger, Schunk, & Toepoel, 2013). For a more detailed description of the panel, the sample, recruitment and response, see the website www.lissdata.nl or Scherpenzeel and Das (2011). All our analyses use unweighted data.

In December 2008 (from here on wave 1) several MTMM questions were administered to 2873 respondents, randomly selected from the panel. Because it is generally not advisable to administer the same question using three different methods in one questionnaire, a Split-ballot version of the MTMM questionnaire was used (Saris, Satorra, & Coenders, 2004). In the LISS panel, every respondent answered only 2 out of the 3 methods on all the traits. Because data are missing by design (Missing Completely At Random), Full Information Maximum Likelihood estimation in Mplus 8.0 (L. Muthén & B. Muthén, 2016) was used in all models to account for the

Table 1

Correlations, means and variances of 9 political trust MTMM questions

Trait	Method	1–1	2–1	3–1	1–2	2–2	3–2	1–3	2–3	3–3
1) Parliament	1) 0-10 battery	1.00	-	-	-	-	-	-	-	-
2) Legal system	1) 0-10 battery	0.77	1.00	-	-	-	-	-	-	-
3) Police	1) 0-10 battery	0.62	0.72	1.00	-	-	-	-	-	-
1) Parliament	2) 0-5 battery	0.74	0.56	0.43	1.00	-	-	-	-	-
2) Legal system	2) 0-5 battery	0.56	0.78	0.56	0.61	1.00	-	-	-	-
3) Police	2) 0-5 battery	0.33	0.43	0.77	0.41	0.54	1.00	-	-	-
1) Parliament	3) 0-10 score	0.85	0.65	0.50	0.76	0.56	0.33	1.00	-	-
2) Legal system	3) 0-10 score	0.71	0.88	0.65	0.56	0.77	0.44	0.71	1.00	-
3) Police	3) 0-10 score	0.54	0.64	0.89	0.43	0.55	0.77	0.52	0.66	1.00
Means		5.64	5.89	6.04	2.88	2.99	3.13	5.67	5.99	6.25
Variance		1.79	1.91	1.75	0.70	1.04	0.98	1.76	1.81	1.61
Sample size		2482	2483	2483	3.55	355	355	370	370	370

missingness. Pre- and post-processing of MPLUS output was done in R 3.3.0. (R Core Development Team, 2016) using the MplusAutomation package (Hallquist & Wiley, 2014).

The MTMM questionnaire contained 9 questions on political trust that were previously asked in the European Social Survey. The 3 traits subsequently ask for trust in 1. The Dutch parliament, 2. The Legal system and 3. The Police. These three traits were combined with three methods: 1.a 0–10 battery, with labeled endpoints, presented horizontally in a battery-type format, 2. a 0–5 battery, again with labeled endpoints and 3. a 0–10 score with labeled endpoints, presented vertically where a respondent writes down the answer. See Appendix B for the full question wordings and response scales. Table 1 shows the correlations, means and variances for the nine variables. The correlation follow the usual structure of MTMM models; they are highest for items that measure the same traits, followed by items using the same methods, and lowest for items that measure different traits and different methods. On all three methods, the mean trust is lowest for the parliament, followed by the legal system and police. Across methods, the mean levels of trust are higher for the 0–10 score than for the 0–10 battery. If the 0–5 battery means were multiplied by two to reflect the range of the 0–10 scale, the means of the 0–5 battery are about the same as the 0–10 score. The variances then also appear highest for the 0–5 battery, followed by the 0–10 battery and the 0–10 score. A more formal analysis of these data is required however to tease out the exact effects of the method and trait factors on the observed statistics. This is what we turn to now.

4 Results

4.1 Cross-sectional MTMM

First, we look at the results from the cross-sectional MTMM model with means. The resulting reliability and validity coefficients are about equal to the scores obtained by Revilla and Saris (2012) despite the slight difference in the specification of our MTMM model as compared to theirs.

Table 2 shows that the questions on ‘trust in parliament’ and ‘trust in the police’ are somewhat more reliable than the question ‘trust in the legal system’. Generally, we see that the validity coefficients are high (>.90) for all 3 methods.

Apart from the reliability and validity coefficients, which show how the variances and covariances are attenuated by random error measurement errors, we also look at bias in the means due to the method being used (systematic errors).

The Latent method means indicates the differences of the second and third method relative to the first method (0–10 item battery). The method means of the second (0–5 battery) and third method (0–10 score) are significantly different from zero, indicating that apart from measurement error affecting the variances and covariances shown earlier in Table 2, there is also relative measurement bias due the method being used (see Table 3). For the 0–5 battery method, this difference reflects the different metric of the scale being used.

4.2 Extension to include attrition

In this section, the cross-sectional MTMM model estimates from the previous section are extended to include attrition. This is done by running a multigroup model 35 times. Each time, the sample is restricted to those respondents, who answer in that particular wave of the LISS panel in the first group, and nonrespondents in the second group. There is considerable wave nonresponse and attrition in the LISS panel, leading to ever smaller realized samples over time. The LISS panel has countered attrition by adding top-up samples to the data, but these are ignored in this paper.

Figure 3 shows attrition in the LISS between December 2008 and December 2011, which are the data used in this paper. In July 2011, no questionnaire was administered, so this wave was left out of our analyses. Attrition in LISS is not

Table 2
Reliability and validity estimates for political trust questions in wave 1

Trait	Method	Reliability $1 - \mathrm{Var}(e)$	Validity λT	Method effect λm
1) parliament	1) 0–10 battery (ref)	0.95	0.97	-
2) legal system	1) 0–10 battery (ref)	0.95	0.82	-
3) police	1) 0–10 battery (ref)	0.95	0.91	-
1) parliament	2) 0–5 battery	0.82	0.97	0.39
2) legal system	2) 0–5 battery	0.86	0.85	0.38
3) police	2) 0–5 battery	0.84	0.93	0.40
1) parliament	3) 0–10 score	0.90	0.97	0.29
2) legal system	3) 0–10 score	0.93	0.83	0.27
3) police	3) 0–10 score	0.94	0.93	0.30

Table 3
Latent Means for political trust questions in wave 1

	Latent Mean α	Std. Err. of the mean	Variance ψ
1) Latent trait mean trust in parliament	5.62	0.03	2.96
2) Latent trait mean trust in legal system	5.87	0.03	3.40
3) Latent trait mean trust in police	6.03	0.03	2.87
1) Latent method mean 0-10 battery (ref)	0.00	0.00	0.00
2) Latent method mean 0-5 battery	−2.77	0.04	0.73
3) Latent method mean 0-10 score	0.15	0.03	0.29

necessarily monotone: respondents can drop out and come back (Lugtig, Das, & Scherpenzeel, 2014). The sample size of 2873 in wave 1 is reduced to 1863 in wave 35.

When running the 35 MTMM models (one for each wave), we found that our model fit the data well in each wave based on values for CFI and RMSEA (see appendix A for full results). In discussing the results from the longitudinal MTMM models we take a similar approach as with the cross-sectional model. We first look at the reliability and validity coefficients of the respondents who remain over time to assess change in measurement error in variances and covariances caused by attrition. The changes we observe are caused by attrition, and thus indicate relative attrition error. As a second step we look how attrition affects the Latent Trait means to assess relative attrition bias over time, and finally, look at change in the Latent Method means that would indicate a relation between measurement bias and attrition.

4.3 Reliability and validity estimates over time

If attrition affects the covariance structure over time, it can do so in two ways. First, the validity coefficients can change over time. If the coefficients become lower over time, this means that those people whose responses are more affected by the method that is used to ask the question (0–10 battery, 0–5 battery or 0–10 score) are dropping out from the study at a higher rate. Similarly, the reliability coefficients may change over time when those respondents with more or less consistent political trust attitudes than average, drop out at a faster rate. In other words, a change over time in the validity and reliability statistics would show that measurement error and attrition interact, and would suggest that the reliability or validity of the political trust questions would change over time.

We find that attrition does not affect any of the validity or reliability coefficients over time. Apart from some small fluctuations shown in Figure 4, there is no upward or downward trend in any of the statistics.

The effects of attrition on means as depicted in figure 5 are similar. As with the covariances, the means are disentangled into a few different parameters: 1. the Latent Trait mean shows the mean of a specific trait using the reference method (0–10 battery) and 2. the Latent Method means show how the mean would change if a different method to ask the trait is used. It is important to keep in mind that all estimates for the means are based on the fact that we use the 0–10 battery as the reference method. So, we assume that the effects of this method are 0 and do not change over time.

The Latent trait means and latent method means are expressed on different metrics. The metric of the latent trait is determined by the original scale we used for the first item in our questionnaire (0–10), while the Latent method mean indicates the difference between the expected mean using the

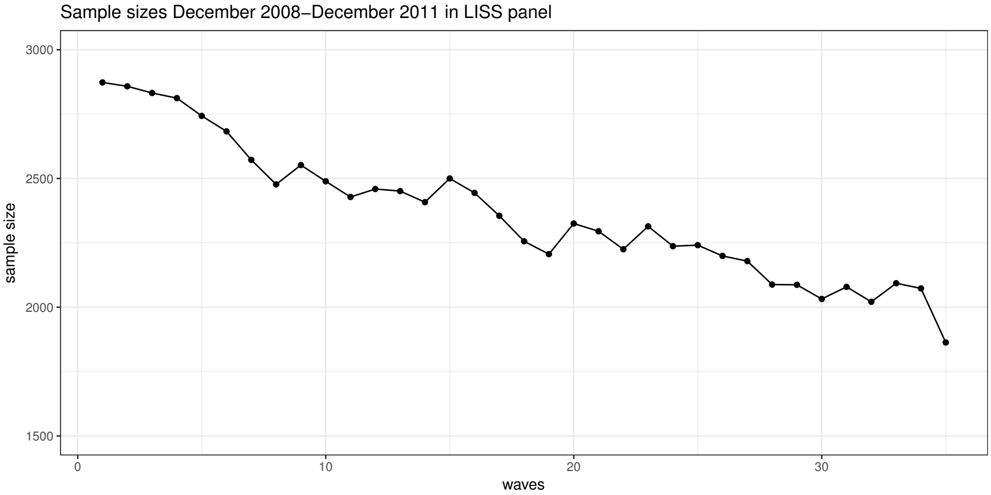

Figure 3. Attrition in the LISS panel between December 2008 and December 2011

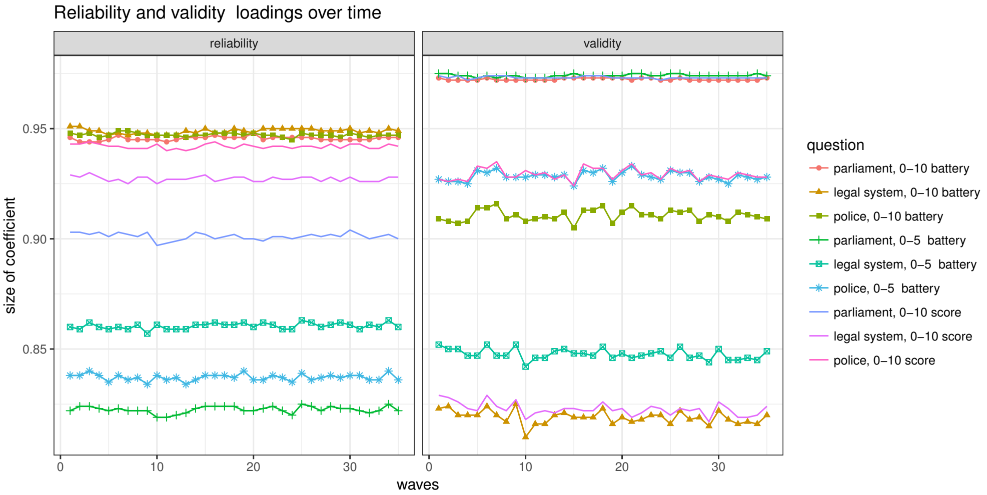

Figure 4. Reliability and validity coefficients over 34 waves of the panel study for remaining respondents

2nd and 3rd question format, compared against the first. In order to be able to compare the Latent Trait and Method means over time, we rescale both variables so that they express the difference between the trait and method means at wave 1(see Table 3) with later waves.

The Latent Trait means for all three political trust questions (1 – parliament, 2- legal system, 3- police) are stable over time. They become a bit higher at first- at most +.04 – implying that people with slightly lower levels of trust drop out. Note that the effects are really small however.

For the Latent method means, we find that the differences between methods 1 (0–10 battery) , and methods 2 (0–5 battery) and 3 (0–10 score) remain stable over time. Please note that although the method means are stable over the long run, there are fluctuations between waves. These fluctuations are however relatively small, as the Latent Method Means never change by more than 0.2 due to attrition.

The conclusion from inspecting the Latent means is thus similar to the response quality. Respondents who drop out of the LISS panel did not report with more measurement error in the MTMM experiment, nor do they have higher or lower political trust (Latent Traits), or do they react differently to the format of the question being offered to them (Latent methods)

5 Discussion

The method presented in this paper to study the relation between attrition and measurement error in panel surveys can easily be used in any panel survey. If an MTMM study is conducted for the key survey variables in one of the first waves of the study, later attrition can be modelled within the MTMM framework, and error and biases due to measure-

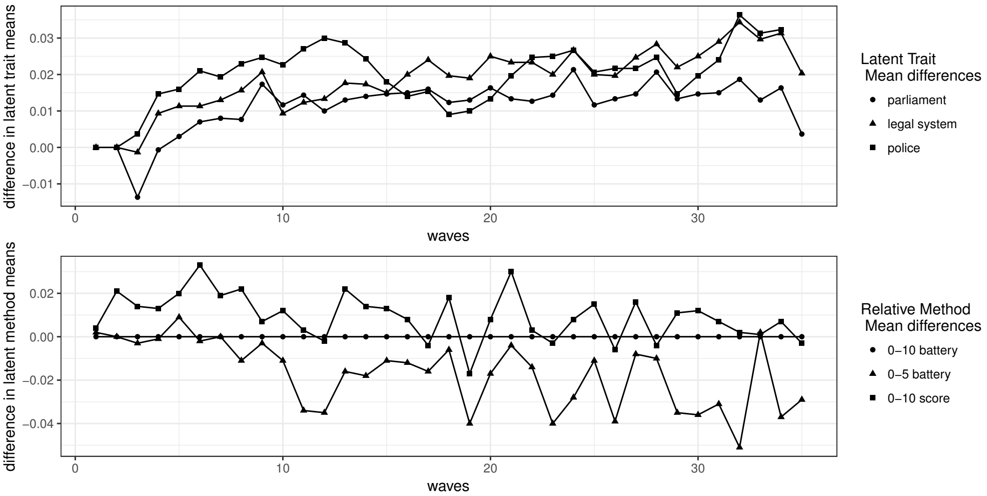

Figure 5. Differences in Latent Trait and Latent method means between wave 1 and the 34 consecutive waves of the LISS panel

ment and attrition can be assessed and expressed on the same metric.

In our example, we found that the reliability and validity of the survey questions on political trust did not change when respondents who later dropped out were excluded from the MTMM-analysis. We found only very small effects for changes in Latent means as well, implying that over time, respondents who report with more measurement error when the MTMM questionnaire was administered were not more likely to drop out later in the panel study. In other words, we found no relation between measurement and attrition error in our data.

The LISS panel did include three more MTMM experiments in the same wave when the political trust MTMM design was administered. These questions focused on 1) media consumption, 2) Life satisfaction and 3) social trust. We repeated our analysis for those variables as well, and found similar results: the reliability and validity coefficients did not change over time, attrition bias was small, and method mean bias was stable over time. One reason why we chose to use political trust as an example in this paper was that the question formats used to measure political trust were relatively comparable. For media consumption, life satisfaction and social trust, the formats differed much more, and because of that, the method variation within those studies was much larger than the variation we found.

One reason for the fact that we find small effects of attrition bias may be that the MTMM questionnaire was only administered in the 12^{th} month of the LISS study. It is likely that much of the attrition error and bias in the LISS study was already introduced by that time. In our study we find that people with lower levels of trust are somewhat more likely to dropout: it is very well possible that this effect disappears, or even changes sign, had the MTMM study been administered earlier. For this reason, future MTMM questionnaire should ideally be administered in the first wave of a panel survey, so that the total bias due to attrition can be compared to measurement error. Arguably, there are other variables that are equally, if not more important to include in the first questionnaire of a panel study, like demographic variables. However, when the panel study focuses largely on attitudinal variables, MTMM questionnaires are crucial to assess not only the extent of measurement error, but also nonresponse error.

MTMM data can be modelled in many different ways. In this paper, we have used a method that yields a relative Method mean bias coefficient, along with the more traditional reliability and validity coefficients. Several other analysis models should also be explored, mainly because they would enable the researcher to study different trade-offs. For example, the effect-coding approach advocated by Pohl and Steyer (2010) allows the Latent Trait and Method factors to be correlated. This can in turn inform survey researchers whether cross-sectionally, respondents with high values on Latent Traits have different method-effects compared to those with low values on the Latent Trait. In order words, whether there is dependence between measurement error and the true value itself. Although the MEref model we used is technically identified when such correlations are added, in practice, our models did not converge, possibly due to the fact that we used a split-ballot design to administer the MTMM questions. Another way to extend our models is to look at different forms of attrition. In this paper, the group of respondents who drop out consist of permanent dropouts, and those who dropout temporarily.

A further topic of future research could study the stability of estimates in MTMM models. Studies by Grimm, Pianta, and Konold (2009) and Koch et al. (2014) are two of the few studies that repeated an entire MTMM questionnaire over time. They find that the MTMM models are measurement invariant over time, and that the parameter estimates are stable. This implies that measurement quality is indeed a stable respondent characteristic, and that because of this an MTMM questionnaire administered at the start of a panel study can be informative for studying the size and relation between attrition and measurement error in later waves. However, we do believe that more frequent administration of MTMM questions in panel studies would still be worthwhile. For example, when top-up samples are added to a panel study, MTMM designs can be used to study the presence of panel conditioning among the existing respondents. Similarly, the effect of mode-switches, now common in panel studies can be investigated in detail.

6 Acknowlegment

The author would like to thank Alex Cernat and Joop Hox for comments on an earlier version of this manuscript. This work was supported by a Future Leaders Grant by the UK Economic and Social Research Council [ES/K001027/1] awarded to the author.

References

Alwin, D. F. (2014). Investigating response errors in survey data. *Sociological Methods & Research*, *43*(1), 3–14.

Biemer, P. P. (2010). Total survey error: design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848.

Binswanger, J., Schunk, D., & Toepoel, V. (2013). Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly*, *77*(3), 783–797.

Callegaro, M. & Disogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*(5), 1008–1032.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *56*(2), 81–105.

Cannell, C. F. & Fowler, F. (1963). Comparison of a self-enumerative procedure and a personal interview: a validity study. *Public Opinion Quarterly*, *27*(2), 250–264.

Coenders, G. & Saris, W. E. (2000). Systematic and random method effects estimating method bias and method variance. *Development in Survey Methodology Metodološki Zvezki*, *15*, 55–74.

Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling*, *11*(4), 535–559.

Couper, M. & Ofstedahl, M. (2009). Keeping in contact with mobile sample members. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 183–203). Chichester: Wiley.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the internet. *Social Science Research*, *38*(1), 1–18.

Felderer, B., Kirchner, A., & Kreuter, F. (2013). *The effect of survey mode on data quality: disentangling nonresponse and response bias*. Presentation at the 24th Workshop on Household Survey Nonresponse. London, 2–4 September.

Fricker, S. & Tourangeau, R. (2010). Examining the relationship between nonresponse propensity and data quality in two national household surveys. *Public Opinion Quarterly*, *74*(5), 934–955.

Grimm, K. J., Pianta, R. C., & Konold, T. (2009). Longitudinal multitrait-multimethod models for developmental research. *Multivariate Behavioral Research*, *44*(2), 233–258.

Groves, R. M. (2005). *Survey errors and survey costs* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Hallquist, M. & Wiley, J. (2014). Package 'mplusautomation'. Version 0.6-3. Retrieved from http://cran.r-project.org/web/packages/MplusAutomation/index.html

Hox, J. J., de Leeuw, E., & Chang, H.-T. (2012). Nonresponse versus measurement error are reluctant respondents worth pursuing? *BMS. Bulletin De Mëthodologie Sociologique*, *113*(1), 5–19.

Kaminska, O., McCutcheon, A., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74*(5), 956–933.

Koch, T., Schultze, M., Eid, M., & Geiser, C. (2014). A longitudinal multilevel CFA-MTMM model for interchangeable and structurally different methods. *Frontiers in Psychology*, *5*, pages.

Kreuter, F., Müller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research: making use of administrative data. *Public Opinion Quarterly*, *74*(5), 880.

Kreuter, F., Müller, G., & Trappmann, M. (2013). A note on mechanisms leading to lower data quality of late or reluctant respondents. *Sociological Methods & Research*, *43*(3), 452–464.

Longitudinal Internet Study for the Social sciences (LISS). (2008). European social survey MTMM (study 23). CentERdata, University of Tilburg. Retrieved from www.lissdata.nl

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Reading, MASS: Addison-Wesley.

Lugtig, P., Das, M., & Scherpenzeel, A. (2014). Nonresponse and attrition in a probability-based online panel for the general population. In M. Callegaro, P. J. Lavrakas, J.

Krosnick, R. P. Baker, J. Bethlehem, & A. S. Göritz (Eds.), *Online panel research: a data quality perspective. Wiley series in survey methodology* (pp. 135–154). New York: Wiley.

Lynn, P. (Ed.). (2009). *Methodology of longitudinal surveys*. Chichester: Wiley.

Lynn, P. & Lugtig, P. (2017). Total survey error for longitudinal surveys. In P. Biemer, E. De Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, . . . B. West (Eds.), *Total survey error in practice.* (Chapter 13). New York: Wiley.

Muthén, L. & Muthén, B. (2016). *Mplus: statistical analysis with latent variables. User's guide (version 8.0)*. Los Angeles, CA: Muthén and Muthén.

Olson, K. (2013). Do non-response follow-ups improve or reduce data quality? A review of the existing literature. *Journal of the Royal Statistical Society. Series A, Statistics in Society*, *176*(1).

Pohl, S. & Steyer, R. (2010). Modeling common traits and method effects in multitrait-multimethod analysis. *Multivariate Behavioral Research*, *45*(1), 45–72.

Pohl, S., Steyer, R., & Kraus, K. (2008). Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society: Series A*, *171*.

R Core Development Team. (2016). *R: a language and environment for statistical computing (version 3.3.0)*. Vienna: R foundation for Statistical Computing.

Revilla, M. A. & Saris, W. E. (2012). A comparison of the quality of questions in a face-to-face and a web survey. *International Journal of Public Opinion Research*, *25*.

Revilla, M. A. & Saris, W. E. (2013). The split-ballot multitrait-multimethod approach: implementation and problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(1), 27–46.

Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly*, *74*(5), 907.

Saris, W. E. & Andrews, F. (1991). Evaluation of measurement instruments using a structural modeling approach. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 575–597). New York: Wiley.

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

Saris, W. E., Satorra, A., & Coenders, G. (2004). A new approach to evaluating the quality of measurement instruments: the split-ballot MTMM design. *Sociological Methodology*, *34*, 311–347.

Scherpenzeel, A. & Das, M. (2011). True longitudinal and probability-based internet panels: evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavorial research and the internet - advances in applied methods and research strategies* (pp. 77–104). New York: Routledge.

Scherpenzeel, A. & Saris, W. E. (1997). The validity and reliability of survey questions. A meta-analysis of MTMM studies. *Sociological Methods & Research*, *25*(3), 341–383.

Watson, N. & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 152–182). Chichester: Wiley.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*(1), 1–26.

Appendix A

Table A1
Model fit statistics for MTMM models extended to attrition

Wave	X^2 Value (df = 53)	CFI	RMSEA	Respondents	Nonrespondents
1	188.69	0.985	0.040	2873	335
2	197.14	0.984	0.041	2858	350
3	171.52	0.987	0.037	2832	376
4	163.40	0.988	0.036	2812	396
5	193.32	0.985	0.041	2743	465
6	159.72	0.988	0.035	2683	525
7	162.15	0.988	0.036	2572	636
8	178.07	0.986	0.038	2477	731
9	162.74	0.988	0.036	2552	656
10	180.27	0.986	0.039	2489	719
11	167.07	0.988	0.037	2428	780
12	172.32	0.987	0.037	2459	749
13	150.06	0.989	0.034	2451	757
14	169.62	0.987	0.037	2408	800
15	180.12	0.986	0.039	2500	708
16	156.93	0.989	0.035	2444	764
17	165.94	0.988	0.036	2355	853
18	161.39	0.988	0.036	2256	952
19	144.64	0.990	0.033	2206	1002
20	180.50	0.986	0.039	2325	883
21	154.69	0.989	0.035	2295	913
22	160.55	0.988	0.036	2225	983
23	169.23	0.987	0.037	2314	894
24	157.41	0.989	0.035	2237	971
25	168.11	0.987	0.037	2241	967
26	189.90	0.985	0.040	2199	1009
27	159.44	0.988	0.035	2179	1029
28	145.13	0.990	0.033	2088	1120
29	162.04	0.988	0.036	2087	1121
30	165.40	0.988	0.036	2032	1176
31	184.05	0.986	0.039	2079	1129
32	171.41	0.987	0.037	2021	1187
33	155.99	0.989	0.035	2093	1115
34	164.26	0.988	0.036	2073	1135
35	177.60	0.986	0.038	1863	1345

RMSEA: Root Mean Square Error of Approximation. CFI: Comparative Fit Index.

Appendix B
Trust in institutions questions

Method 1

Please indicate on a score of 0–10 how much you personally trust of the institutions listed below. 0 means you do not trust an institution at all, and 10 means you have complete trust.

- *Method 1/Trait 1:* The Dutch parliament?

0 No trust at all	1	2	3	4	5	6	7	8	9	10 Complete trust	11 Don't know
□	□	□	□	□	□	□	□	□	□	□	□

- *Method 1/Trait 2:* The Legal system?

0 No trust at all	1	2	3	4	5	6	7	8	9	10 Complete trust	11 Don't know
□	□	□	□	□	□	□	□	□	□	□	□

- *Method 1/Trait 3:* The police?

0 No trust at all	1	2	3	4	5	6	7	8	9	10 Complete trust	11 Don't know
□	□	□	□	□	□	□	□	□	□	□	□

Method 2

Please select one box below to show how much you personally trust each institution.

- *Method 2/Trait 1:* The Dutch parliament?

0 No trust at all	1	2	3	4	5 Complete trust	6 Don't know
□	□	□	□	□	□	□

- *Method 2/Trait 2:* The Legal system?

0 No trust at all	1	2	3	4	5 Complete trust	6 Don't know
□	□	□	□	□	□	□

- *Method 2/Trait 3:* The police?

0 No trust at all	1	2	3	4	5 Complete trust	6 Don't know
□	□	□	□	□	□	□

Method 3

Method 3/Trait 1. Please indicate on a scale of 0 to 10 how much you personally trust the Dutch parliament. If you have no trust at all give a score of 0. If you have complete trust, give a score of 10, otherwise give a number in between.

- Score: 0 . . . 10
- don't know

Method 3/Trait 2. Please indicate on a scale of 0 to 10 how much you personally trust the legal system. If you have no trust at all give a score of 0. If you have complete trust, give a score of 10, otherwise give a number in between.

- Score: 0 . . . 10
- don't know

Method 3/Trait 3. Please indicate on a scale of 0 to 10 how much you personally trust the police. If you have no trust at all give a score of 0. If you have complete trust, give a score of 10, otherwise give a number in between.

- Score: 0 . . . 10
- don't know