Political Science Research and Methods (2026), **14**, 472–489
doi:10.1017/psrm.2025.31

ORIGINAL ARTICLE

Measurement error when surveying issue positions: a MultiTrait MultiError approach

Kim Backström[1], Alexandru Cernat[2], Rasmus Sirén[1] and Peter Söderlund[3]

[1]Social Science Research Institute, Samforsk, Åbo Akademi University, Abo, Finland; [2]Department of Social Statistics, University of Manchester, Manchester, UK and [3]Swedish School of Social Science, University of Helsinki, Helsinki, Finland
Corresponding author: Kim Backström; Email: kim.backstrom@abo.fi

(Received 13 December 2023; revised 8 October 2024; accepted 27 November 2024)

Abstract
Voters' issue preferences are key determinants of vote choice, making it essential to reduce measurement error in responses to issue questions in surveys. This study uses a MultiTrait MultiError approach to assess the data quality of issue questions by separating four sources of variation: trait, acquiescence, method, and random error. The questions generally achieved moderate data quality, with 76% on average representing valid variance. Random error made up the largest proportion of error (23%). Error due to method and acquiescence was small. We found that 5-point scales are generally better than 11-point scales, while answers by respondents with lower political sophistication achieved lower data quality. The findings indicate a need to focus on decreasing random error when studying issue positions.

Keywords: latent variable modelling; measurement error; MTME; issue positions; survey methodology

1. Introduction

Spatial models of voting, which highlight the importance of politicians' and voters' issue positions, are central in political science. The parties and individual voters are distributed across the policy space—or rather, policy spaces—since elections are often contested on multiple relevant policy dimensions. The spatial thesis posits that voters prefer parties and candidates whose policy positions closely align with their own (Downs, 1957; Laver, 2014). However, accurately measuring voters' positions on specific issues requires valid and reliable tools. The most common method for assessing individual voter positions is through batteries of attitude questions in mass opinion surveys.

There are no gold standard measurements for preferences, making evaluating their data quality difficult. Measurement error can be affected by aspects such as question formulation, response options, and scale levels (e.g., Krosnick, 1991). Previous research has shown that data quality for survey questions, especially nonfactual ones, can be low (Alwin, 2021). Furthermore, using the European Social Survey (ESS), Saris et al. (2011) estimated the average measurement quality to be 0.64, while Alwin (2007) found the reliability of survey questions in a U.S. context to be as low as 0.5. Also, measurement error can be both random and correlated. The types and sizes of measurement errors create different amounts of bias and lead to varying strategies for correction (Saris and Revilla, 2016; Cernat and Oberski, 2022). However, previous research on measurement error rarely estimated multiple sources of measurement error concurrently (Cernat and Oberski, 2019). Finally, measurement error can also be moderated by respondent characteristics (Cernat and Toepoel, 2022). This means that differences

© The Author(s), 2025. Published by Cambridge University Press on behalf of EPS Academic Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

in survey responses to attitudinal questions used when approximating issue positions can, in reality, result from differences in measurement error, for example, between respondents with higher or lower political sophistication.

It is imperative to decrease and correct measurement errors to avoid drawing conclusions about democratic well-being based on flawed data. This paper investigates how measurement error impacts survey items used when studying issue preferences. We implement a novel MultiTrait MultiError (MTME) experimental design in two waves of the Finnish online panel Citizens' Opinion. The MTME design enables concurrently estimating measurement error due to acquiescence, method effects, and random error, while reducing carryover effects. Furthermore, due to the potential moderating effect of political sophistication, we also investigate the effects of political interest, internal political efficacy, and whether respondents have a degree on measurement error. Using this approach, we contribute to a better understanding of the data quality of questions used when studying issue positions, and the findings can be used when designing future studies and to correct for measurement error.

2. Issue positions

A substantial body of literature shows that citizens' preferences over public policy issues significantly influence various outcomes. Classic spatial theories of elections state that individuals assess candidates and parties based on their positions on issues (see Ansolabehere et al., 2008). The influence of public policy preferences on voting decisions is commonly referred to as "policy voting" or "issue voting," with the terms often used interchangeably. The policy voting model demonstrates considerable explanatory power in models of vote choice (Wagner and Kritzinger, 2012; Kessenich and van der Brug, 2024), although some scholars are skeptical about the extent to which issues influence vote choice (see Guntermann and Persson, 2023). The model traces its origins to Downs (1957), who posited that political preferences could be conceptualized along a single left–right axis and citizens opt for the party that gives them the highest utility. Also, when applied to political issues broadly, the model suggests that demand and supply must align whereby citizens prefer the alternative that best represents their policy positions (Ansolabehere et al., 2008; Wagner and Kritzinger, 2012).

In empirical studies of citizen preferences, respondents are asked to locate themselves on various policy or issue scales. Individuals may have preferences on many different matters. Such preferences involve a wide range of concrete topics such as taxation, redistribution of wealth, privatization, environmental protection, law and order, defense spending, and immigration. Studies addressing such preferences vary in scope, from using single survey items for specific topics to multiple survey items for different policy dimensions. The dimensionality of policy spaces implies that distinct bundles of issue positions are correlated. Data-reduction techniques map respondents' positions on specific issues to latent policy dimensions, such as left–right economic policy and liberal–conservative social policy (Laver, 2014). Survey items on political attitudes are often combined to measure more general political values such as economic left–right and libertarian–authoritarian orientations (Kumlin and Knutsen, 2005). However, it is important to clearly distinguish political values from issues, as the former are defined as "prescriptive beliefs about which goals one would like to see implemented in the political system and about the desired participatory forms to influence politics," while the latter are "often more narrowly defined—capturing particular policy proposals or political circumstances" (Aardal and Van Wijnen, 2005: 195).

Scales composed of multiple measures, by averaging several items together or constructing factor scores, indeed improve our ability to assess people's underlying predispositions that are coherent and stable (Ansolabehere et al., 2008). However, irrespective of whether we use individual survey items to tap preferences over public policy issues or issue scales composed of multiple measures, we should improve the quality of individual survey items to reduce measurement error. Different types of survey items in terms of response alternatives and their order are used. A scan of the literature shows that

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

studies gauging issue preferences have used 5-point (Hellwig, 2014), 7-point (Dinas et al., 2016), or 11-point scales (Isotalo et al., 2019). Other studies using data from the European Values Study or the ESS have combined responses on 2-, 3-, 4-, and 10-point scales (Knutsen, 2018) or five- and 11-point scales (van der Brug and Van Spanje, 2009). There is also variation in how the survey questions are framed, for example, whether the government should take measures to reduce differences in income levels (those who agree are economically left) and whether there should be a lower taxation level (those who agree are economically right).

3. Measurement error

The concept of measurement error can be understood as the difference between the theoretical concept of interest and the collected data. Measurement error and other errors such as coverage, sampling, and nonresponse errors contribute to a difference between the measured sample statistic and the population parameter according to the total survey error framework (Groves and Lyberg, 2010). Considering our study design, we will focus on three types of measurement error: method effects, acquiescence, and random error.

Method effects include respondents' tendency to answer questions in a specific manner regardless of the substance of the question. In survey research, these are often conceptualized as the impact of the response scale (Andrews, 1984; Saris and Gallhofer, 2014), including aspects such as the number of scale levels and whether numeric or verbal labels are used. The method effects are typically studied using the MultiTrait MultiMethod (MTMM) approach. For example, Saris and Gallhofer (2014) show that method effects varied between 0.36 and 0.50 when studying three questions across three forms in an MTMM experiment embedded in the ESS.

The main takeaway from other research on method effects is that fewer scale points are preferable when studying preferences (Revilla et al., 2014; Alwin et al., 2018; Höhne et al., 2023). However, DeCastellarnau (2018) found in a literature review that aspects such as scale length and using verbal and numeric labels affect data quality, and the results are mixed due to different aspects being measured and interaction effects. Given the contradicting results and the interaction of scale characteristics, there is a need for experimental designs that separate different causes of measurement error concurrently.

Acquiescence is the tendency to agree with statements regardless of content (Krosnick, 1991; Billiet and McClendon, 2000). In a survey methodology context, this tendency can be influenced by how a question or statement is worded and how the response options are designed. When using bipolar survey questions with response options ranging from agree to disagree, a survey item with the first response option "agree" indicates a possible increased risk of satisficing. This type of measurement error can also be present when using labels such as "good," "average," and "bad" (Hofmans et al., 2005). Furthermore, the occurrence and strength of the acquiescence appear to vary based on the topic and survey design (Hofmans et al., 2007; DeCastellarnau, 2018; Keusch and Yan, 2018). Previous MTME studies looking at acquiescence and other error sources concurrently have found acquiescence variance to vary across forms and questions but generally to be minor compared to other error sources (Cernat and Oberski, 2019, 2022, 2023).

Random error describes the "noise" when respondents answer survey questions (Alwin, 2007). Respondents interpret questions differently, may not have established opinions or attitudes related to the question, or may not understand the question, leading to random response patterns that vary with every new measure. However, as opposed to the two previously discussed correlated errors, method effects and acquiescence, random error consists of differences in response error specific to a single question (Cernat and Oberski, 2022). However, while random error does not bias estimates of the population mean due to its randomness, it does bias regression coefficients (Fuller, 1987), and can inflate estimates of change when using longitudinal data (Cernat and Sakshaug, 2021). Simply ignoring this "noise" due to its random nature is not valid when striving to make accurate inferences.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Random error can also be conceptualized as the complement of reliability. Alwin (2021) found, looking at three General Social Survey panel studies, that reliability varied around 0.85 when looking at factual survey questions and around 0.66 when looking at nonfactual survey questions. However, reliability could be as low as 0.52 for some nonfactual questions. Furthermore, even when accounting for other error sources concurrently in other MTME studies looking at attitudes toward immigrants, random error has, in every case, been the largest source of measurement error (Cernat and Oberski, 2019, 2022, 2023).

Finally, some respondents can be expected to generate more measurement error. Considering the political topic, we may expect more survey errors from respondents with lower **political sophistication**. Agreeing with a statement and choosing the first presented response option decreases the cognitive burden for the respondent while having many response options increases it. The burden may be higher for respondents who do not have crystallized views on policy issues, leading to an increased risk of acquiescence and method effects among respondents with lower interest in the survey topic and lower cognitive ability (Cernat and Toepoel, 2022). Furthermore, Groves (2004) found that non-attitudes are more likely among respondents with lower education and political interest, while the risk of not comprehending the question is higher. Furthermore, when studying attitude stability, Freeder et al. (2019) found that those more politically sophisticated had more stable attitudes when looking at more extended periods. Based on this, respondents with lower political sophistication can also be expected to generate more random error since they are likelier to not have as grounded attitudes as other respondents, especially if the questions used suffer from poor design, such as being overly complex or deemed irrelevant to the respondents.

The literature review highlights that measurement error is perversive in survey research, especially when measuring nonfactual topics. Accounting for this is essential as they can severely impact research results. For example, Saris and Revilla (2016) show that when the quality of two variables decreases, the correlation between the two variables will decrease much faster, leading to an underestimation of the variables' relationship if measurement error is not corrected in the analysis. Also, considering the previous discussion on the potential moderating effects of political sophistication, it is vital to study measurement error among those with lower political sophistication. A better understanding of measurement error allows us to correct for measurement error by designing better survey questions and using statistical models. This, in turn, enables us to produce better quality research that can be used when developing theory and policy.

To improve our understanding of measurement error when studying issue positions, we will investigate the following:

- RQ1: How do acquiescence, method effects, and random error affect the measurement quality of issue positions?
- RQ2: How does political sophistication moderate measurement error?

4. Methodology

4.1. The MTME framework

The MTME is a generalization of within-person experimental designs, such as test–retest or the MTMM (Cernat and Oberski, 2019, 2022). A within-person experimental design implies asking respondents the same questions multiple times. The design makes estimating reliability possible by examining how consistently respondents answer the same question. The simple design can also be expanded to include correlated sources of bias. For example, MTMM designs, in addition to asking the same questions, also manipulate the response scale of the questions, with respondents receiving different answer options in the follow-up (Campbell and Fiske, 1959; Saris and Andrews, 1991). This makes it possible to estimate "methods effects," which is a source of correlated variance due to the response scale and not the content of the question.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

The MTME framework generalizes this approach and implies the development of experimental manipulations that can lead to different sources of measurement error. Furthermore, the MTME framework is theory-driven and allows researchers to study relevant errors and estimate their relative impact. For example, Cernat and Oberski (2022) manipulated the response scale, the categories' order, and the questions' content to estimate method effects, acquiescence, and social desirability (in addition to random error). Such a complex design leads to multiple "forms" of the questions (way to word the question and response category). A split ballot factorial design is typically used to minimize the burden on the respondents (Saris et al., 2004; Revilla et al., 2019), in which respondents are randomly allocated to groups and receive the questions twice using different forms.

The MTME framework offers a flexible way to experimentally explore different measurement error sources. That being said, the approach is a within-person experimental design and thus inherits its limitations, such as memory effects (people remembering they were asked the questions already). Possible solutions to this issue are increasing the time between re-interviews (although too long of a time might lead to true change in the topic of interest), randomly changing the order of the forms, or accounting for memory effects (e.g., controlling for cognitive ability or period to reinterview). In the next sections, we will discuss how we developed an MTME design to estimate measurement error in responses to issue questions, how we modeled the data, and how we minimized and controlled for memory effects.

4.2. The MTME design and data

The MTME experiment design was included in waves five ($n = 3{,}763$; RR = 75.2%) and six ($n = 3{,}885$; RR = 76.9%) of the Finnish panel Citizens' Opinion (Finnish: Kansalaismielipide) 2023 parliamentary election study (Grönlund and Strandberg, 2023). The panel is recruited via nonprobability and probability sampling, and data are collected online using Qualtrics. The two waves used in this study were conducted in the period from April 4 to 24, 2023, after the parliamentary election on the 2nd of April. The entire election study consisted of 4,875 respondents, while this dataset only uses respondents who answered the relevant questions in both waves ($n = 3{,}175$).[1] See Backström et al. (2023) for a more in-depth description of the panel and the election study. Data for the entire election study can be accessed at the Finnish Social Science Data Archive (FIRIPO and Strandberg, 2023), while data and replication files for the MTME experiment can be accessed at the PSRM Harvard Dataverse (Backström et al., 2025).

Ten survey items were included in the two waves using a bipolar matrix question that asked respondents: 'What is your opinion on the following political proposals?' The items are used to capture different issue dimensions prevalent in Finland, such as the traditional economic left–right dimension, and various aspects of the sociocultural dimension. We also include an item used to capture the historically relevant center-periphery dimension. Westinen (2015) presents an overview of the political cleavage structures in Finland, and Söderlund (2023) discusses a more in-depth analysis of the substantive results and the ability of the items to capture different dimensions in the election study. The matrix questions were included at the end of the respective waves, and the total wave response times were about 10 minutes (wave five median: 10.5 minutes; wave six median: 10.1 minutes). Table 1 presents the exact statement wordings and what political issue dimension the items attempt to capture.

The experimental design is based on varying the number and direction of scale points that respondents can choose from when answering the questions. These variations are done to capture measurement errors resulting from method effects (scale points) and acquiescence (scale direction) while estimating random error. The survey attributes were manipulated in a 2 × 2 experimental

[1] When studying dropout due to not answering both waves, we find that there are significant differences between respondents included versus those not included. There is an overrepresentation of men, older, higher educated, non-probability recruited, politically interested, and with high internal political efficacy.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Table 1. Survey items used and their representative issue dimension

Item number	Statement-wording	Issue dimension
1	**Tax** solutions that reduce income differences in society	Economic
2	Allowing private **corporations** to be responsible for elder care to a greater degree	Economic
3	Increasing financial support to **rural** areas	Center-Periphery
4	Implementing stricter **prison** sentences	Sociocultural
5	Accepting fewer **refugees** to Finland	Sociocultural
6	Strengthening **gay**, bi- and transsexual rights in society	Sociocultural
7	Investing more in an **environment**ally friendly society	Sociocultural
8	Finland should leave the **EU**	Sociocultural
9	Increasing the **digitalization** of society	Other
10	Introducing a climate tax on **beef**	Sociocultural

Table 2. Question forms when varying scale points and directions (2 × 2)

Form number	Number of scale points	Scale direction
F1	5	Very good proposal → Very bad proposal
F2	5	Very bad proposal → Very good proposal
F3	11	Very good proposal → Very bad proposal
F4	11	Very bad proposal → Very good proposal

design, where respondents could answer using either 5- or 11-point scales, and the scale direction was either from good to bad proposal (decremental) or bad to good proposal (incremental). All scale points had verbal labels in the 5-point scale forms,[2] while only the polar scale points had verbal labels in the 11-point forms. Furthermore, all scale points in the 11-point forms had numeric labels ranging from 0 to 10. These variations resulted in four possible forms for the question (Table 2).

However, the MTME approach is based on a within-person experimental design, meaning that respondents must answer two forms of the question at two different points to estimate measurement error. Using a split ballot design (Saris and Gallhofer, 2014), the four forms over two measurement points result in six combinations, or form pairs, that must be administered to achieve the design.[3] However, the form pairs' order should be randomized to counteract potential carryover effects. Randomizing the order of the forms results in 12 treatment groups (2 × 6) (Table 3).

Memory effects are also a concern when carrying out within-person experimental designs. Respondents may remember they answered the same questions earlier, affecting their answers. For example, they may simply repeat the previously given answer due to satisficing (Rettig and Blom, 2021). Increasing the time between measurement points can decrease the risk of memory effects. Previous research has found that memory effects are rare after at least 20 minutes of interview time (Saris et al., 2010; Saris, 2013). However, new research found respondents to be able to reproduce their answers after 20 minutes across questions on beliefs, attitudes, and behavior (Rettig et al., 2023; Rettig and Struminskaya, 2023), indicating a need for more time between measurement points.

However, when studying measurement error using within-person experimental designs, keeping time between measurement points too long can also be problematic since there is a risk of true change in the concepts of interest (Cernat and Oberski, 2019). This means that differences observed between

[2]For example, F1: very good proposal, fairly good proposal, neither good nor bad proposal, fairly bad proposal, very bad proposal. Every form also had a "don't know" response option, these responses have been recoded as NA. See Figure A1 in the appendix for the proportion of item nonresponse.

[3]$4!/((2!(4-2)!)) = 6$.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Table 3. Randomized form order combinations across measurement points for the treatment groups ($n = 3{,}175$)[4]

Group number	T1 form	T2 from	Group sample size
1	F1	F2	265
2	F1	F3	285
3	F1	F4	299
4	F2	F3	270
5	F2	F4	259
6	F3	F4	251
7	F2	F1	251
8	F3	F1	275
9	F4	F1	259
10	F3	F2	258
11	F4	F2	256
12	F4	F3	247

forms of the questions for the same respondent could be caused by a change in the underlying trait and not by measurement error. However, preferences on sociocultural issues generally are more stable (Zaller, 2012), meaning that survey items used should not be prone to true change in traits.

The two waves that included the MTMEexperiment took place after the parliamentary election, meaning there should not be an election outcome effect that could have changed the true score between the two waves. However, the waves' field periods overlapped each other in time. To ensure the correct form orders and decrease the impact of memory effects, we excluded respondents who answered the waves in the wrong order ($n = 10$) or both waves on the same day ($n = 46$). This resulted in 3,175 respondents, with a mean time between waves of 9.6 days (min: 1; max: 19; sd: 2.3). Ensuring there was at least 1 day between responses, randomizing the form order, the fact that both waves occurred after the election, and the relative stability of the topic of interest should balance well the possible confounders of memory effects and true change. We will also run sensitivity analyses to investigate how the time between measurements impacts error estimates.

Furthermore, from a data quality standpoint, probability samples are preferable to nonprobability samples since analyzing data from nonprobability samples increases requirements regarding transparency, design, and modeling assumptions (Baker et al., 2013). The critique on nonprobability samples is based on the uncertainty regarding being able to make accurate inferences about a larger population. Lately, this critique has focused on online incentivized opt-in panels such as Amazon's Mechanical Turk (Kennedy et al., 2020; Ahler et al., 2021). However, the Citizens' Opinion panel does not employ comparable incentives for panel participants, and almost all of the panelists recruited via nonprobability sampling have been in the panel for over 3 years, meaning they are active and willing participants even without incentives used in other online panels. Still, while there are exceptions (e.g., Einarsson et al., 2022), based on the previous literature, nonprobability samples can be expected to generate more measurement errors. Because of this, we will decompose variance between panelists recruited via non-probability ($n = 1{,}154$) and probability ($n = 2{,}021$) sampling as a sensitivity analysis.

[4] A post hoc power analysis showed that a sample size of 3,175 with 1 numerator degree of freedom, 12 groups, 0 covariates, an alpha of 0.05, and an effect size of 0.1 would result in a power of 0.9998802. The experimental groups do not differ significantly with respect to gender and education, while there are some differences between age-groups (see Table A1 in the appendix).

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

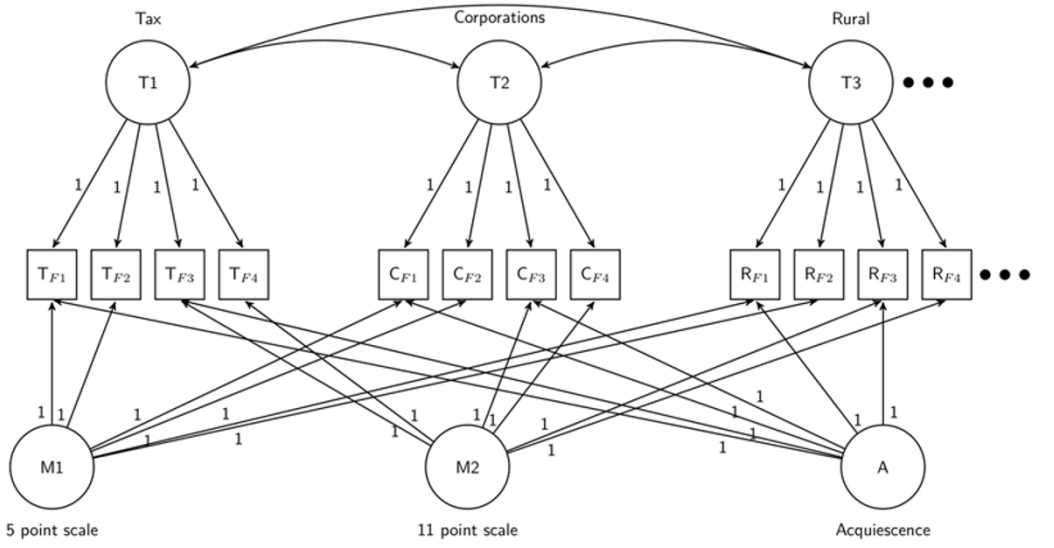


Figure 1. Representation of the MTME model estimated in the SEM framework. Observed variables are represented by squares, each topic being measured using four different forms (F1–F4). Latent variables, or factors, are represented by circles. The "T" latent variables represent the concept of interest, while the "M" latent variables represent method effects due to the response scale, and "A" represents acquiescence caused by the direction of the scale. Only 3 out of the 10 topics are presented for ease of reading. Residual errors are not presented for the same reason.

4.3. The MTME model

Based on the experimental design presented in the previous section, we develop a statistical model that separates three different sources of measurement error from the trait, or concept of interest (Cernat and Oberski, 2019, 2022). First, we estimate **method effects** as the variance due to the response scale, 5 versus 11 points. Second, we estimate **acquiescence effects**, or the tendency to agree with statements regardless of their content, by comparing situations when positive response categories are presented first versus when presented last. This is based on the expectation that it is easier to agree with a statement if the positive response category is presented first. Finally, we also estimate **random errors** or noise.

To separate these different sources of variation concurrently, we develop an MTME model in the Structural Equation Modeling framework (Bollen, 1989):

$$y^*_{tma} = \lambda^{(T)^*}_{tma} T_t + \lambda^{(M)^*}_{tma} M + \lambda^{(A)^*}_{tma} A + \varepsilon_{tma}$$

where y^*_{tma} is the observed variable measuring a particular trait or topic, t, a method or response scale, m, and a scale direction, a. We decompose these observed variances in four sources of variation: T, measuring the **trait**, M, measuring the **method** effect, A, measuring **acquiescence** and an item specific **random error**, ε_{tma}. The trait variance represents the valid source of variation that measures the concept of interest. We reason that, even when working with items used to create indices, it is essential to detect measurement error at the individual-item level to achieve the best possible data quality. The Acquiescence and Method variances are correlated measurement errors as they represent consistent answering patterns due to the format of the response scale and not the content. The random error represents noise in the data that can bias confidence intervals and multivariate analyses. The visual representation of the model can be seen in Figure 1.

In the MTME models, correlated measurement error (such as Method (M) or Acquiescence (A)) can be identified using either a latent variable for each condition or using a dummy coding approach

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

(Cernat and Oberski, 2019, 2022). Here, we use the former approach for the method effect, thus estimating two latent variables (*M1* and *M2*), one for each method used, and the latter for acquiescence (*A*). As a result, the method effects can be interpreted as the amount of variation due to each method compared to a hypothetical answer without a method effect. The acquiescence latent variable variance can be interpreted as the amount of extra variation due to presenting a positive response category first compared to presenting it last. Using the dummy coding approach can help estimate the models (as used in the MTMM-1 models; see Eid et al., 2008) but can sometimes be harder to interpret.[5]

In addition to running the MTME model for the entire sample, we also investigate if MTME estimates vary by key groups identified in the literature. We will do this by separating the MTME estimates by the degree of political interest, degree of internal political efficacy, and education level, which we view as indicators of political sophistication.[6]

Political interest is indicative of general interest in answering the survey, considering the political theme. Groves (2004) found an association between survey topic interest and degree of non-attitude and acquiescence, meaning respondents with lower political interest can be expected to achieve lower data quality on questions used to measure political issue positions. Political interest is estimated through the question "Generally, how interested are you in politics?" with the response options "very interested," "'fairly interested," "not particularly interested," and "not interested at all." However, to achieve better class balance, the response options "fairly interested," "not particularly interested," and "not interested at all" were combined in the analyses, creating two groups that either have "high political interest" ($n = 1{,}275$) or "low political interest" ($n = 1{,}888$).

Internal political efficacy is seen as an individual's assessment of their ability to understand what is going on in politics (Niemi et al., 1991). Considering that a lack of knowledge or comprehension of a subject indicates non-attitudes (Groves, 2004), it is reasonable to assume that a low level of internal political efficacy is also associated with non-attitudes on political issues. A higher propensity to answer these surveys randomly will lead to a greater overall degree of error. We use three survey items and the four response options to create a sum variable (scale 0-9), where a higher value indicates a higher degree of internal political efficacy. See Table A2 in the appendix for a closer decomposition of the items and response options used to create the sum variable and differences within the sample regarding internal political efficacy. To achieve class balance, we recoded the sum variable into a dichotomous variable where respondents with a value under the mean (6.28) were coded as having a low degree of internal efficacy ($n = 1{,}438$), while those over the mean were coded as having a high degree ($n = 1{,}238$).

Finally, having an academic degree is also viewed as indicative of political sophistication and cognitive ability. Again, being less educated is associated with a higher chance of non-attitude and acquiescence (Groves, 2004). Meisenberg and Williams (2008) also found that cognitive ability or education level is negatively related to acquiescence and extreme response styles. Having a degree was measured using the question "What is your highest achieved education?" with nine response options. The responses were then recoded based on whether the respondent had achieved a university or university of applied sciences degree ($n = 1{,}666$) or not ($n = 1{,}489$).

We view these variables as indicators of political sophistication, meaning a respondent with low political interest, low internal political efficacy, and/or without an academic degree can be expected to have lower political sophistication compared to those with a higher formal sophistication concerning politics. Political interest, internal political efficacy, and education level are all closely related and can be used as proxies for political knowledge (Rapeli, 2022). Furthermore, high political knowledge

[5]For the models where there were estimation issues, we have run a model where both M and A are identified using the dummy approach method. These are mentioned in Section 5.

[6]See Table A2 in the appendix for the descriptive statistics.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

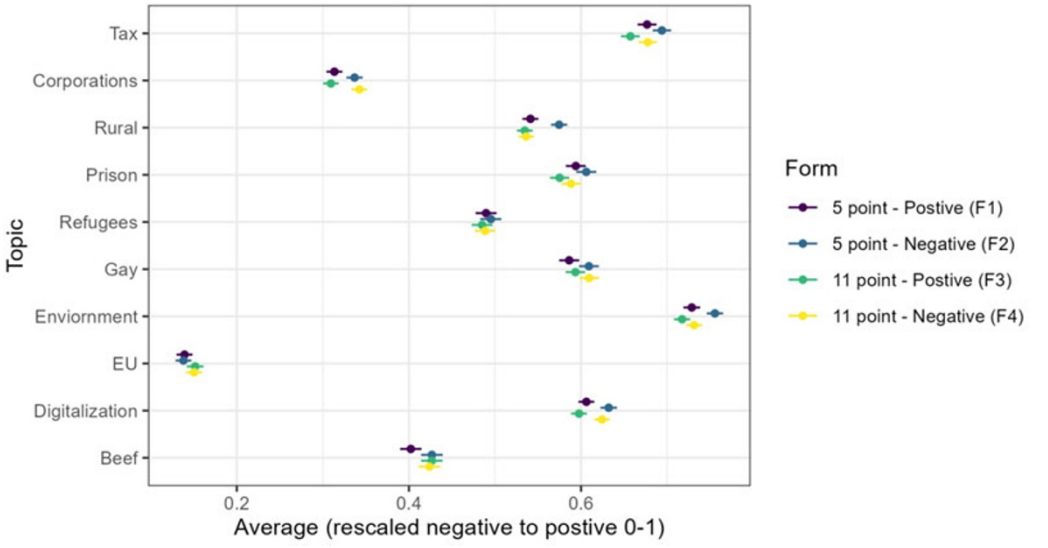


Figure 2. Rescaled averages and confidence intervals by form and topic.

is related to correct voting (Lau et al., 2013; Pierce and Lau, 2019), meaning those politically knowledgeable can better vote according to their preferences. Based on this, we expect respondents with low political sophistication to be more error-prone.

Data were cleaned in R 4.3.1 (R Core Team, 2023), and the SEM models were estimated in the lavaan package (Rosseel, 2012). Missing data[7] were dealt with using Full Information Maximum Likelihood, which assumes Missing At Random (MAR) given the model (Enders, 2022).

5. Results

We first explore the results using descriptive statistics. To investigate if the response scale (5 vs. 11 points) or the scale direction (positive vs. negative first) impact mean estimates, we rescaled all the questions on a 0–1 scale, and we reverse coded the positive first forms (Figure 2) so larger numbers imply more support for the statement. Overall, the averages are similar, and the response scale used does not seem to bias mean estimates. Some notable exceptions are the Corporations questions, where there is a significant difference between positive first and negative first. Surprisingly, people are less likely to select the first category when this is positive. A similar pattern can be observed for the 5-point scale for Rural, Environment, and Digitalization.

Next, we create a correlation matrix of all the questions we use in the analysis (4 forms × 10 topics). In line with MTMM research, we expect the four forms measuring the same concept to be highly correlated (Campbell and Fiske, 1959). Ideally, the correlation of different topics using the same forms should be smaller (implying low method effects). Such a pattern would imply that the content of the question is the main driver of the answers and not the form of the response scale. Our correlation matrix (Figure 3) presents such a pattern, with the lighter squares on the off-diagonal representing higher correlations for questions measuring the same topic. Based on these, we expect relatively low method and acquiescence effects in the MTME model and relatively high data quality.

[7]Item nonresponse (NA and "Don't know"-answers) proportions varied between 6.9% (F1, environment) and 12.2% (F3, corporations). See Figure A1 in the appendix for a decomposition.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

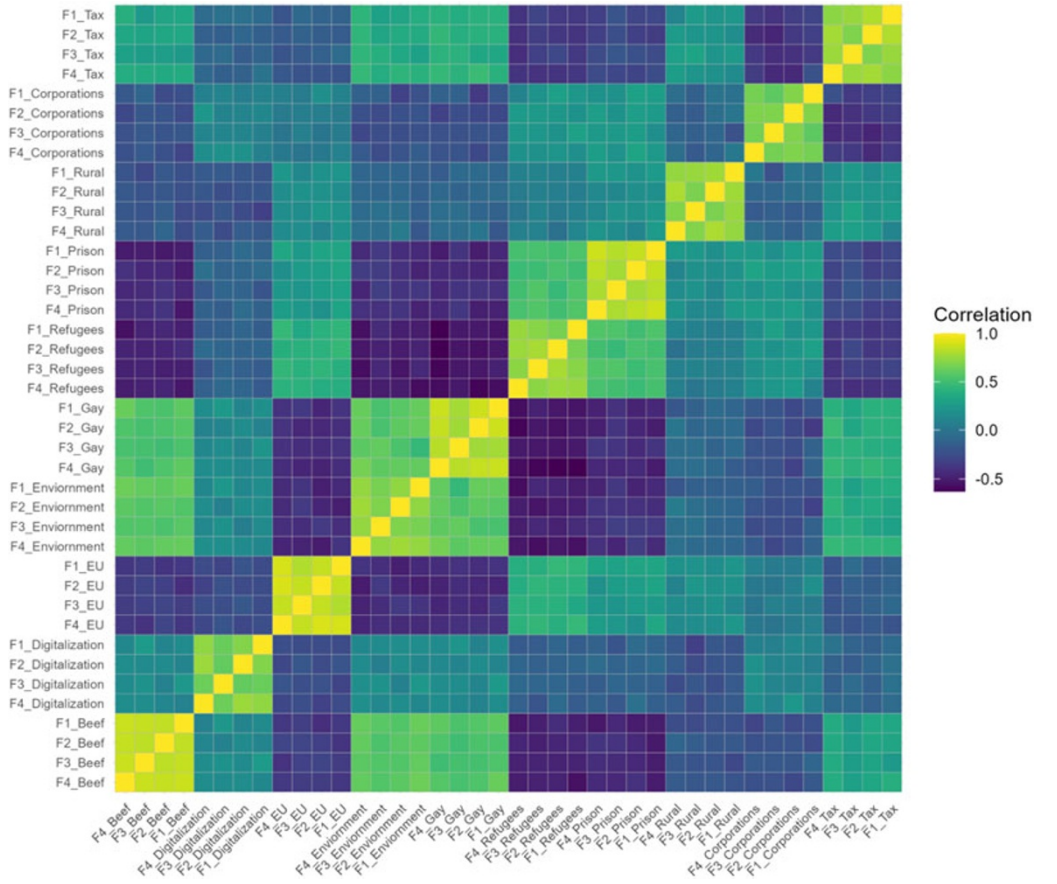


Figure 3. Correlation matrix of all the questions and forms.

5.1. MTME results

The model proposed in Section 4.3 was estimated successfully and had an overall good fit (χ^2: 1535.617, *df*: 719, *p*-value: 0.000, CFI: 0.978, RMSEA: 0.019, SRMR: 0.026). Using the model, we can decompose the total amount of observed variance in the sources presented in Section 4: trait, acquiescence, method, and random error. Figure 4 presents the variance decomposition by response scale and topic.

Overall, the 10 questions measuring issue positions show moderate data quality. The proportion of variance measuring the concept of interest (trait, i.e., validity) is approximately 76%. Looking at the causes of measurement error, it appears that random error is the largest source, with around 23% variance. By comparison, method and acquiescence variance is small, representing less than 1%. The data quality is similar when using either 5-point or 11-point response scales (trait variance 78% vs. 75%). When using 11-point response categories, we have slightly more random error (24% vs. 21%) and method effects (1.2% vs. 0.2%). Looking at data quality by topic, we see considerable variation. Four topics have trait variance over 80% (higher quality): EU, Beef, Gay, and Prison, while two have lower than 70%: Corporations and Digitalization.

We can use this approach to investigate further which response scale leads to the best data quality for each topic or question. Figure 5 highlights that in most cases, the 5-point scale leads to better

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

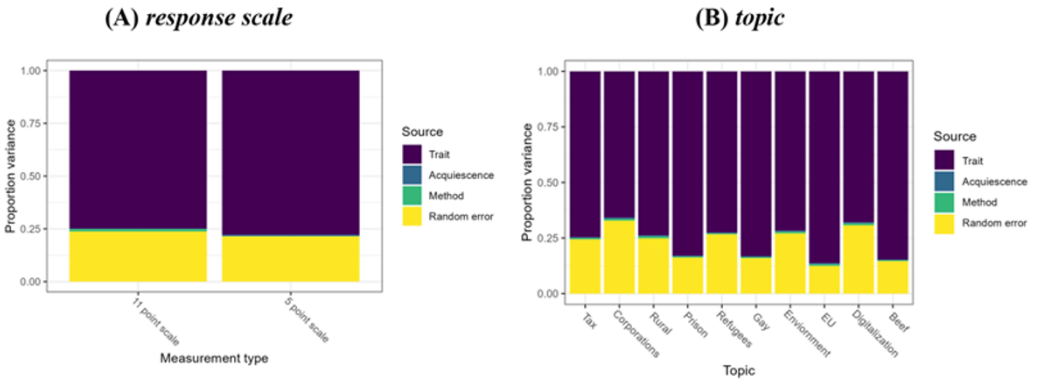


Figure 4. Variance decomposition by (a) response scale and (b) topic.

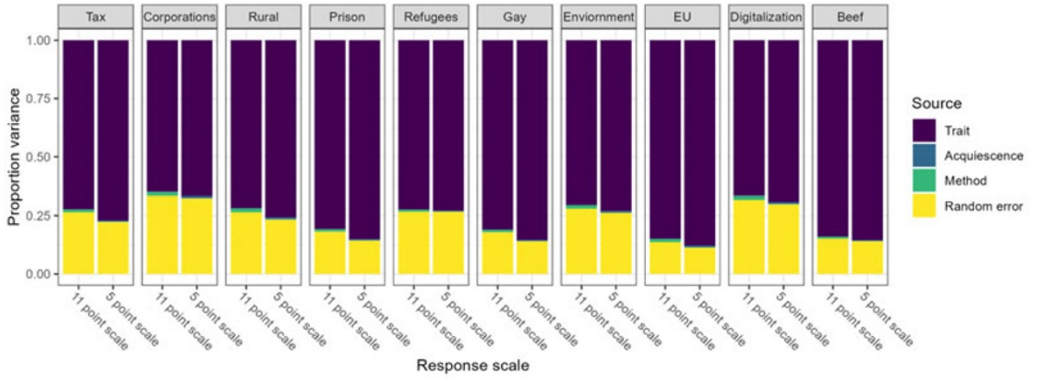


Figure 5. Variance decomposition based on MTME by topic and response scale.

quality compared to the 11-point. The only question where the response scale does not seem to matter is Refugees.

5.2. Moderating effects

These overall data quality patterns could hide important group differences (Cernat and Toepoel, 2022). As a result, we investigate if the data quality of 10 items measuring issue positions is moderated by other factors such as political interest, internal political efficacy, and degree. To answer this research question, we run a series of multigroup MTME models where the moderating factors define the groups. This allows each group to have different data quality indicators.

First, comparing data quality for those with high interest to those with low interest in politics, we observe slightly higher quality for the latter.[8] Overall, those with high political interest have around 78% trait variance versus 76% of those with low interest. The difference is caused by more random error for the latter (24% vs. 22%) and acquiescence effect (1% vs. 0.3%). We see considerable variation when comparing the two groups by topic (Figure 6). For example, the Corporations question has a difference of approximately 12% in data quality (73% vs. 60%), with those with less interest in politics showing lower quality. The other questions show smaller differences, and for the EU question, those with lower interest actually show slightly higher quality (88% vs. 86%).

[8]Note that for this model, we use dummy coding both for the method and acquiescence factors (similar to mtmm-1) to help with the estimation of the model and avoid negative variances (Haywood cases). See Section 4 for a discussion regarding this.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

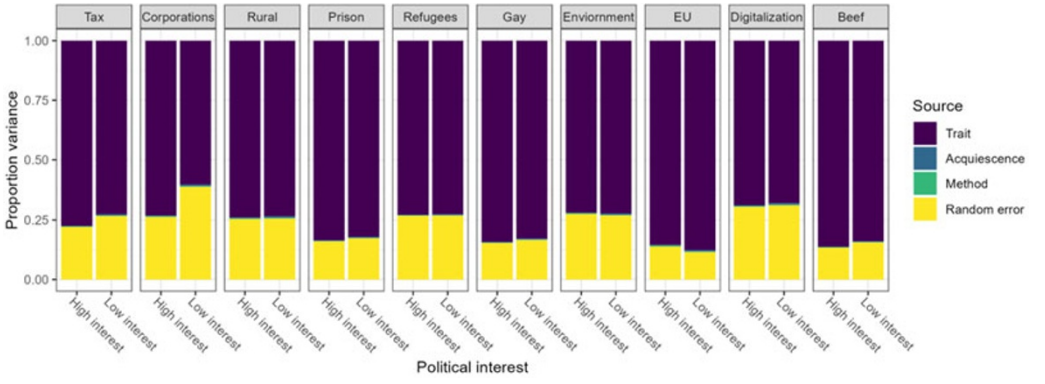


Figure 6. Variance decomposition based on a multigroup MTME by topic and political interest.

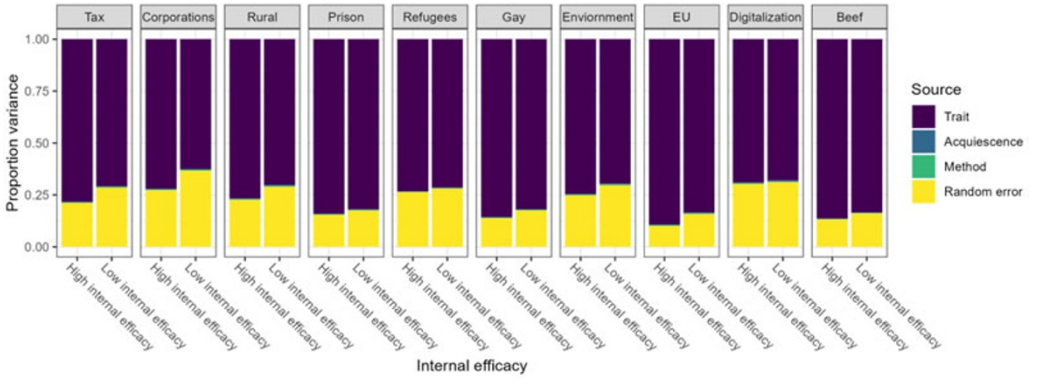


Figure 7. Variance decomposition based on a multigroup MTME by topic and internal efficacy.

Similar patterns are observed when comparing those with high internal efficacy with those with lower levels.[9] Overall, trait variance is lower for those with low internal efficacy (79% vs. 74%). This seems to be caused mainly by more random error for the latter group (25% vs. 21%). Looking at the different topics (Figure 7), we observe the largest differences for Corporations (10%), Tax (8%), Rural (7%), EU (6%), and Environment (5%).

Finally, we also investigate if having a degree leads to differences in measurement errors.[10] Overall, we observe more trait variance for those with a degree (77% vs. 75%). Again, this seems to be mainly driven by random error (22% vs. 25%). This pattern can also be observed for most topics (Figure 8). The largest differences can be observed for Tax (11%) and Corporations (8%), with those without a degree having less trait variance and more random error.

5.3. Sensitivity analysis

We also run two sensitivity analyses. First, we investigate whether the time between interviews impacts data quality estimates. We do this by rerunning the MTME model while controlling for the

[9]Note that for this model, we use dummy coding both for the method and acquiescence factors (similar to mtmm-1) to help with the estimation of the model and avoid negative variances (Haywood cases). See Section 4 for a discussion regarding this.

[10]Note that for this model, we use dummy coding both for the method and acquiescence factors (similar to mtmm-1) to help with the estimation of the model and avoid negative variances (Haywood cases). See Section 4 for a discussion regarding this.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

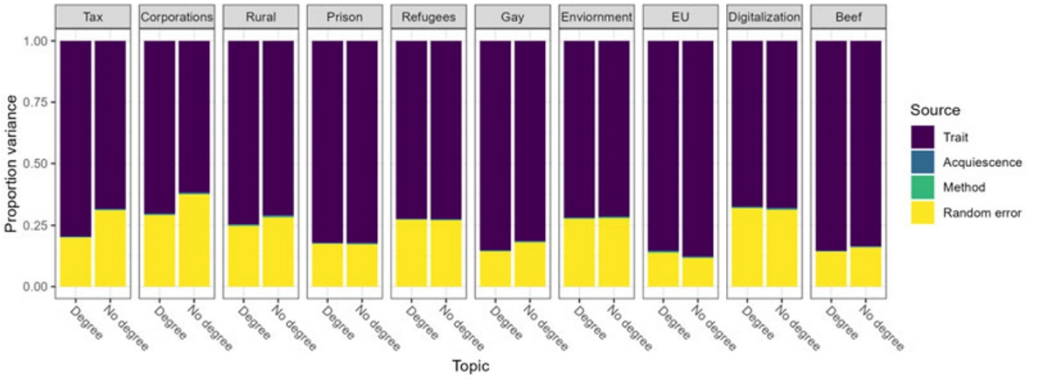


Figure 8. Variance decomposition based on a multigroup MTME by topic and degree.

number of days between the interviews. We do not observe any differences in our estimates after controlling for the number of days (Figure A2 in the appendix).

We also investigate if data quality is different by sampling source. More precisely, we compare those selected through random sampling and those self-selected in the study. Overall, the differences between the probability and nonprobability samples are relatively small. For the trait variance, this is only 0.8% (76.1% for probability vs. 76.9%). We observe more variation by looking at each question separately (Figure A3 in the appendix). The largest difference is for the Corporations variable, where the non-probability sample shows better quality (70% vs. 64% trait variance). The second largest difference is in the opposite direction for the Digitalization variable (65% vs. 70%).

6. Discussion

In this paper, we investigated the quality of 10 items used when measuring positions on political issues in Finland. We used an experimental design, the MTME, to examine how acquiescence, method effects, and random error affect these important measures. Overall, the items have moderate data quality, with around 76% of the variance measuring the topic of interest (trait variance), 23% being random error, and around 0.7% for method and acquiescence. The data quality varied significantly by topic, with the Corporations and Digitalization questions having the lowest quality (under 70% trait variance) and EU, Beef, Gay, and Prison showing the highest quality (over 80% trait variance).

While method and acquiescence had a low impact on the overall variance, the choice of response scale has proved important. The data quality is slightly better for the 5-point compared to the 11-point response scales (trait variance 78% vs. 75%). This is true for most topics investigated here. Looking at average estimates, there are few differences, but when differences appeared, surprisingly, people were less likely to select the first category when this is positive.

We also investigated if data quality varies by key groups. We looked at how political interest, internal political efficacy, and having a degree moderate data quality. While the differences were small, we observed a systematic pattern with people with less interest, less internal efficacy, and, without a degree, having lower data quality. This seems to be driven mainly by random error. The differences are especially large for some topics, such as Tax and Corporations, which are used to tap economic left–right preferences. These might be topics where some of the public does not have enough information to have informed opinions.

As with all research, this also has some limitations. While we used an experimental design to investigate data quality, our data have some potential limitations. First, memory effects could impact the results since our design had a reinterview. To minimize this, we restricted our data to people with at least 1 day between reinterviews. We have also randomized the order in which respondents

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

were presented with the response forms, minimizing potential systematic biases. We find the same results when running a sensitivity analysis by controlling for the days between interviews (Figure A2 in the appendix). Additionally, the sample comprises a mix of respondents who were selected using probability and non-probability sampling. We observed minor differences when analyzing the two groups separately (Figure A3 in the appendix). Still, the interview and re-interview design led to us being unable to include respondents who only answered one of the survey waves, leading to differential dropout. Considering the differences between included and excluded respondents concerning political sophistication, measurement error can be even higher in non-filtered survey data. Also, it is possible that measurement error would be higher in a less labor-intensive cross-sectional survey, as opposed to this study's panel setting. A panel might have filtered away those with lower political sophistication (especially interest) that otherwise would have answered a cross-sectional survey.

In addition, while our implementation of the MTME offers the possibility to separate three sources of variance (method, acquiescence, and random error), other causes, such as social desirability, could be present. Future research should investigate to what degree other sources of measurement error are present when measuring issue preferences. Similarly, here we investigated just 5- and 11-point response scales, but future research could compare them with other alternatives, varying the number of response scale points, verbal and numeric scales, and item-specific questions.

Considering the study's inherent limitations, we can nevertheless draw some important conclusions. First, survey items estimating issue positions have moderate data quality and are similar to other opinion or attitudinal questions. For example, Alwin (2021) has shown by analyzing hundreds of questions from multiple longitudinal studies that factual questions collected in surveys have a reliability of around 0.85, while nonfactual items often have reliabilities as low as 0.5. Saris et al. (2011) estimated the average measurement quality at 0.64 using 2,460 questions from multiple cross-sectional studies. We also note that the main threat to data quality seems to be random error. In contrast, systemic errors due to methods and acquiescence are small, which aligns with previous MTME studies that looked at different error sources concurrently (Cernat and Oberski, 2019, 2022, 2023). Random error can be problematic, as Saris and Revilla (2016) show that a true correlation of 0.9 can appear to be 0.33 when the quality coefficient of the two questions used is 0.6. They further show how regression coefficients and their relative importance can change due to measurement error.

In addition to this overall moderate data quality, we note that this varies by key population groups, political interest, internal efficacy, and degree. This can be especially problematic as differences observed in issue positions between these groups could be caused by measurement error, not differences in the concept of interest. This can have important implications for political theory and policy.

Researchers can account for this information by designing better data collections that minimize measurement errors in advance and correcting for errors after data collection. We recommend using 5-point response scales instead of 11-point ones as they show better quality. Similarly, presenting the negative response category first may minimize acquiescence effects. Also, we recommend avoiding the Tax, Corporations, and Digitalization questions or developing alternatives, as these have shown the largest differences in quality across key groups or overall low quality.

Researchers should also consider correcting for measurement error. Since the main threat stems from random error, researchers should focus on that. Researchers, for example, could re-interview a random sample of the data to estimate reliability (e.g, using a test–retest approach). Those reliability estimates can be used to correct for measurement error in standard statistical packages or using a SEM framework. Researchers can also consider developing their own MTME or MTMM designs to estimate data quality. The advantage of these models is that the trait latent variables can be used in substantive analysis or saved for other users. In this way, they would be correcting for measurement error in their analyses.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Supplementary material. To view supplementary material for this article, https://doi.org/10.1017/psrm.2025.31. To obtain replication material for this article, please visit https://doi.org/10.7910/DVN/LFFOX1.

Acknowledgements. We thank the editor and reviewers for their valuable feedback on the manuscript. We are also grateful to participants at the American Association for Public Opinion Research Annual Conference (2024) and the Nordic Political Science Congress (2024) for their insightful comments and suggestions for future research.

Funding statement. This research received no specific funding.

Competing interest. The authors declare none.

Data availability statement. Data and replication files are available in the PSRM Dataverse (Backström et al., 2025).

References

Aardal B and Van Wijnen P (2005) Issue Voting. In Thomassen J (ed), *The European Voter: A Comparative Study of Modern Democracies*. Oxford: Oxford University Press, 192–212.

Ahler DJ, Roush CE and Sood G (2021) The Micro-Task Market for Lemons: Data Quality on Amazon's Mechanical Turk. In *Political Science Research and Methods*. Cambridge University Press. doi:10.1017/psrm.2021.57

Alwin DF (2007) *Margins of Error: A Study of Reliability in Survey Measurement*. New Jersey: John Wiley & Sons.

Alwin DF (2021) Developing Reliable Measures—An Approach to Evaluating the Quality of Survey Measurement Using Longitudinal Designs. In Cernat A and Sakshaug JW (eds), *Measurement Error in Longitudinal Data*. Oxford: Oxford University Press, 113–154.

Alwin DF, Baumgartner EM and Beattie BA (2018) Number of Response Categories and Reliability in Attitude Measurement. *Journal of Survey Statistics and Methodology* **6**(2), 212–239. doi:10.1093/jssam/smx025

Andrews FM (1984) Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. *The Public Opinion Quarterly* **48**(2), 409–442. doi:10.1086/268840

Ansolabehere S, Rodden J and Snyder JM (2008) The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Political Science Review* **102**(2), 215–232. doi:10.1017/S0003055408080210

Backström K, Cernat A, Siren R and Söderlund P (2025) Replication Data for: Measurement Error When Surveying Issue Positions: A MultiTrait MultiError Approach. Harvard Dataverse, V1. doi:10.7910/DVN/LFFOX1,

Backström K, Grönlund K and Strandberg K (2023) "Technical Appendix." In *Finland Turned Right: Voting and Public Opinion in the Parliamentary Election of 2023*, edited by Grönlund K and Strandberg K, 1st ed., 137–140. Samforsk, The Social Science Research Institute, Åbo Akademi University. https://urn.fi/URN:ISBN:978-952-12-4300-4.

Baker R, Michael Brick J, Bates NA, Battaglia M, Couper MP, Dever JA, Gile KJ and Tourangeau R (2013) Summary Report of the AAPOR Task Force on Non-Probability Sampling. *Journal of Survey Statistics and Methodology* **1**(2), 90–105. doi:10.1093/jssam/smt008

Billiet JB and McClendon MJ (2000) Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling* **7**(4), 608–628. doi:10.1207/S15328007SEM0704_5

Bollen KA (1989) *Structural Equations with Latent Variables*. New York: John Wiley and Sons.

Campbell DT and Fiske DW (1959) Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin* **56**(2), 81–105. doi:10.1037/h0046016

Cernat A, and Oberski DL (2019) *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment* Extending the Within-Persons Experimental Design: The Multitrait-Multierror (MTME) Approach. In Lavrakas PJ, Traugott MW, Kennedy C, Holbrook AL and de Leeuw E (eds). John Wiley & Sons.

Cernat A and Oberski DL (2022) Estimating Stochastic Survey Response Errors Using the Multitrait-Multierror Model. *Journal of the Royal Statistical Society. Series A: Statistics in Society* **185**(1), 134–155. doi:10.1111/rssa.12733

Cernat A and Oberski DL (2023) Estimating Measurement Error in Longitudinal Data Using the Longitudinal MultiTrait MultiError Approach. *Structural Equation Modeling* **30**(4), 592–603. doi:10.1080/10705511.2022.2145961

Cernat A, and Sakshaug JW (2021) *Measurement Error in Longitudinal Data*. Oxford: Oxford University Press.

Cernat A and Toepoel V (2022) How Do Social and Economic Status Impact Measurement Error? *International Journal of Social Research Methodology*. doi:10.1080/13645579.2022.2122227

DeCastellarnau A (2018) A Classification of Response Scale Characteristics That Affect Data Quality: A Literature Review. *Quality and Quantity* **52**(4), 1523–1559. doi:10.1007/s11135-017-0533-4

Dinas E, Hartman E and van Spanje J (2016) Dead Man Walking: The Affective Roots of Issue Proximity Between Voters and Parties. *Political Behavior* **38**(3), 659–687. doi:10.1007/s11109-016-9331-2

Downs A (1957) *An Economic Theory of Democracy*. Harper: New York.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Eid M, Nussbeck FW, Geiser C, Cole DA, Gollwitzer M and Lischetzke T (2008) Structural Equation Modeling of Multitrait–Multimethod Data: Different Models for Different Types of Methods. *Psychological Methods* **13**(3), 230–253. doi:10.1037/a0013219

Einarsson H, Sakshaug JW, Cernat A, Cornesse C and Blom AG (2022) Measurement Equivalence in Probability and Nonprobability Online Panels. *International Journal of Market Research* **64**(4), 484–505. doi:10.1177/14707853221085206

Enders CK (2022) *Applied Missing Data Analysis*, 2nd edn. Guilford Press.

FIRIPO KG and Strandberg K (2023) "Kansalaismielipide: Eduskuntavaalikyselyt 2023 [Sähköinen Tietoaineisto]. Versio 1.0 (2023-08-22)." Yhteiskuntatieteellinen Tietoarkisto. 2023. http://urn.fi/urn:nbn:fi:fsd:T-FSD3789.

Freeder S, Lenz GS and Turney S (2019) The Importance of Knowing 'What Goes with What': Reinterpreting the Evidence on Policy Attitude Stability. *Journal of Politics* **81**(1), 274–290. doi:10.1086/700005

Fuller WA (1987) *Measurement Error Models*. John Wiley & Sons: New York.

Grönlund K and Strandberg K (2023) *Finland Turned Right: Voting and Public Opinion in the Parliamentary Election of 2023*. Åbo: Samforsk, The Social Science Research Institute, Åbo Akademi University. https://urn.fi/URN:ISBN:978-952-12-4300-4.

Groves RM (2004) *Survey Errors and Survey Costs*. New Jersey: John Wiley & Sons Inc.

Groves RM and Lyberg L (2010) Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* **74**(5), 849–879. doi:10.1093/poq/nfq065

Guntermann E and Persson M (2023) Issue Voting and Government Responsiveness to Policy Preferences. *Political Behavior* **45**(2), 561–584. doi:10.1007/s11109-021-09716-8

Hellwig T (2014) The Structure of Issue Voting in Postindustrial Democracies. *Sociological Quarterly* **55**(4), 596–624. doi:10.1111/tsq.12072

Hofmans J, Cools W, Verbeke P, Verresen N and Theuns P (2005) "A Study on the Impact of Instruction on the Response Strategies Evoked in Participants." In *Proceedings of the Twenty-First Annual Meeting of the International Society for Psychophysics*, edited by John SM and Sonya M. S, 119–132. Traverse City, Michigan: The International Society for Psychophysics. https://proceedings.fechnerday.com/index.php?journal=proceedings&page=issue&op=view&path%5B%5D=15.

Hofmans J, Theuns P, Baekelandt S, Mairesse O, Schillewaert N and Cools W (2007) Bias and Changes in Perceived Intensity of Verbal Qualifiers Effected by Scale Orientation Leuven Gent Management School Ghent. *Survey Research Methods* **1**, http://www.surveymethods.org

Höhne JK, Krebs D and Kühnel S-M (2023) Investigating Direction Effects in Rating Scales with Five and Seven Points in a Probability-Based Online Panel. *Survey Research Methods* **17**(2), 193–204. doi:10.18148/srm/2023.v17i2.8006

Isotalo V, Söderlund P and von Schoultz Å (2019) Polarisoituuko Politiikka Suomessa? Puolueiden Äänestäjäkuntienarvosiirtymät 2003–2019. *Politiikan Ilmastonmuutos: Eduskuntavaalitutkimus 2019*. Helsinki: Oikeusministeriö, 288–306. https://www.researchgate.net/publication/344490567

Kennedy R, Clifford S, Burleigh T, Waggoner PD, Jewell R and Winter NJG (2020) The Shape of and Solutions to the MTurk Quality Crisis. *Political Science Research and Methods* **8**(4), 614–629. doi:10.1017/psrm.2020.6

Kessenich E and van der Brug W (2024) New Parties in a Crowded Electoral Space: The (In)stability of Radical Right Voters in the Netherlands. *Acta Politica* **59**(3), 536–556. doi:10.1057/s41269-022-00269-0

Keusch F and Yan T (2018) Is Satisficing Responsible for Response Order Effects in Rating Scale Questions? *Survey Research Methods* **12**(3), 259–270. doi:10.18148/srm/2018.v12i3.7263

Knutsen O (2018) *Social Structure, Value Orientations and Party Choice in Western Europe*. Palgrave Macmillan: London.

Krosnick JA (1991) Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* **5**(3), 213–236. doi:10.1002/acp.2350050305

Kumlin S and Knutsen O (2005) Value Orientations and Party Choice. In Thomassen J ((ed)), *The European Voter: A Comparative Study of Modern Democracies*. Oxford: Oxford University Press, 125–166.

Lau RR, Patel P, Fahmy DF, and Kaufman RR (2013) Correct Voting across Thirty-Three Democracies: A Preliminary Analysis. *British Journal of Political Science* **44**(2), 239–259. doi:10.1017/S0007123412000610

Laver M (2014) Measuring Policy Positions in Political Space. *Annual Review of Political Science* **17**, 207–223. doi:10.1146/annurev-polisci-061413-041905

Meisenberg G and Williams A (2008) Are Acquiescent and Extreme Response Styles Related to Low Intelligence and Education? *Personality and Individual Differences* **44**(7), 1539–1550. doi:10.1016/j.paid.2008.01.010

Niemi R, Craig S and Mattei F (1991) Measuring Internal Political Efficacy in the 1988 National Election Study. *The American Political Science Review* **85**(4), 1407–1413. doi:10.2307/1963953

Pierce DR and Lau RR (2019) "Polarization and Correct Voting in U.S. Presidential Elections." *Electoral Studies* **60** (August). doi:10.1016/j.electstud.2019.102048.

Rapeli L (2022) "What Is the Best Proxy for Political Knowledge in Surveys?" *PLoS ONE* **17** (8 August). doi:10.1371/journal.pone.0272530.

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: https://www.R-project.org/.

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Rettig T and Blom AG (2021) Memory Effects as a Source of Bias in Repeated Survey Measurement. In Cernat A and Sakshaug JW (eds), *Measurement Error in Longitudinal Data*. 1st Oxford University Press, 3–18.
Rettig T, Blom AG and Karem Höhne J (2023) Memory Effects: A Comparison Across Question Types. *Survey Research Methods* **17**(1), 37–50. doi:10.18148/srm/2023.v17i1.7903
Rettig T and Struminskaya B (2023) Memory Effects in Online Panel Surveys: Investigating Respondents' Ability to Recall Responses from a Previous Panel Wave. *Survey Research Methods* **17**(3), 301–322. doi:10.18148/srm/2023.v17i3.7991
Revilla MA, Saris WE and Krosnick JA (2014) Choosing the Number of Categories in Agree-Disagree Scales. *Sociological Methods & Research* **43**(1), 73–97. doi:10.1177/0049124113509605
Revilla M, Bosch OJ and Weber W (2019) Unbalanced 3-Group Split-Ballot Multitrait–Multimethod Design? *Structural Equation Modeling* **26**(3), 437–447. doi:10.1080/10705511.2018.1536860
Rosseel Y (2012) Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48**(2), 1–36. doi:10.18637/jss.v048.i02
Saris W (2013) Is There Anything Wrong with the MTMM Approach to Question Evaluation? *Pan- Pacific Management Review* **16**(1), 47–77.
Saris WE and Andrews FM (1991) "Evaluation of Measurement Instruments Using a Structural Modeling Approach." In *Measurement Errors in Surveys*, edited by Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA and Sudman S, 575–597. New York: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section. doi:10.1002/9781118150382.ch28.
Saris WE and Gallhofer I (2014) *Design, Evaluation, and Analysis of Questionnaires for Survey Research*, 2nd edn. New York: John WIley & Sons.
Saris WE, Oberski D, Revilla M, Zavala D, Lilleoja L, Gallhofer I and Gruner T. 2011. "The Development of the Program SQP 2.0 for the Prediction of the Quality of Survey Questions." 24. RECSM Working Paper. Barcelona.
Saris WE and Revilla M (2016) Correction for Measurement Errors in Survey Research: Necessary and Possible. *Social Indicators Research* **127**(3), 1005–1020. doi:10.1007/s11205-015-1002-x
Saris WE, Revilla M, Krosnick JA and Shaeffer EM (2010) Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options. *Survey Research Methods* **4**, http://www.surveymethods.org
Saris WE, Satorra A and Coenders G(2004) A New Approach to Evaluating the Quality of Measurement Instruments: The Split-Ballot MTMM Design. *Sociological Methodology* (December):311–347. doi:10.1111/j.0081-1750.2004.00155.x.
Söderlund P. 2023. "Political Value Orientations." In *Finland Turned Right: Voting and Public Opinion in the Parliamentary Election of 2023*, edited by Grönlund K and Strandberg K, 49–56. https://urn.fi/URN:ISBN:978-952-12-4300-4.
van der Brug W and Van Spanje J (2009) Immigration, Europe and the 'new' Cultural Dimension. *European Journal of Political Research* **48**(3), 309–334. doi:10.1111/j.1475-6765.2009.00841.x
Wagner M and Kritzinger S (2012) Ideological dimensions and vote choice: Age group differences in Austria. *Electoral Studies* **31**(2), 285–296. doi:10.1016/j.electstud.2011.11.008
Westinen J (2015) *Cleavages in Contemporary Finland - A Study on Party-Voter Ties*. Åbo Akademi University. https://urn.fi/URN:ISBN:978-951-765-801-0
Zaller J (2012) What Nature and Origins Leaves Out. *Critical Review* **24**(4), 569–642. doi:10.1080/08913811.2012.807648

https://doi.org/10.1017/psrm.2025.31 Published online by Cambridge University Press

Cite this article: Backström K, Cernat A, Sirén R and Söderlund P (2026) Measurement error when surveying issue positions: a MultiTrait MultiError approach. *Political Science Research and Methods* **14**, 472–489. https://doi.org/10.1017/psrm.2025.31