Survey Research Methods (2018)
Vol. 12, No. 3, pp. 193-209
doi:10.18148/srm/2018.v12i3.7207

ISSN 1864-3361
http://www.surveymethods.org

A Closer Look at Attitude Scales with Positive and Negative Items. Response Latency Perspectives on Measurement Quality

Jochen Mayerl
Faculty of Behavioural and Social Sciences
Chemnitz University of Technology, Germany

Christoph Giehl
Faculty of Social Sciences
University of Kaiserslautern, Germany

Measurement quality in standardized surveys has been a core issue for decades in survey research. For questionnaire designers, it is common to use a mixture of positive and negative worded items to measure multi-item-constructs in order to control for response effects like acquiescence bias. The paper shows that paradata such as response latency measurement can be used to identify specific subgroups of respondents with specific types of cognitive response modes. These response modes moderate the occurrence of response effects, systematic and random measurement errors, and thus the reliability and validity of attitudinal measurement models. Therefore, by adapting paradata to detect low measurement quality, it can be used as a tool that leads to a better understanding of respondents' cognitive processes. Data from a German CATI-survey, with experimental design and measurement of response latencies, are used to analyze data quality of a measurement model of attitudes towards health nutrition with mixed items. Response effects are analyzed through the experimental variation of the question order of negative and positive worded items. Structural equation models are estimated in a multiple-group moderator design to test validity and reliability of the latent attitude construct. As a result, the attitude scale shows acceptable values of validity and reliability only under the condition of spontaneous answers where question order effects appear.

Keywords: Response Latency; Paradata; Reversed Items; Method Effects; Cognition in Surveys

1 Introduction

Evaluating the quality of survey measurements and developing methods to avoid or control for measurement errors has been a core issue in survey research for decades. Biasing effects, which lead to survey responses that do not reflect the "true value", are often called response effects. Well-known response effects, which appear independent of question content, are response sets like acquiescence bias, extreme responding, or a tendency to the middle category. In addition, content specific biases like social desirability or question order effects are another source of loss in the measurement quality of surveys. All these response effects lead to systematic biases in survey responses and thus are a serious problem of validity of measurements.

For survey designers, it is quite common to mix positively and negatively worded items to operationalize multiple-item constructs for the purpose of reducing perceived redundancy and controlling for acquiescence. However, mixing item directions may come along with some serious methodological problems, leading to low correlations between positively and negatively worded items and multidimensional instead of unidimensional constructs.

In our paper, we seek to better understand the problems of multi-item attitude scales with positive and negative items by using a special type of paradata: response latency measurements. The response latencies will be used as proxy measures of cognitive processes occurring while respondents answer survey questions in a CATI-study. Our research question is whether this method can help to identify subsamples of respondents with high versus low validity and reliability in the measurement of a multi-item attitude construct that contains mixed item-wording directions.

First, we introduce the methodology of reversed items in multiple-item scales (section 2), followed by a brief overview on paradata with a focus on response latency measurement in survey research (section 3). A dual-mode cognitive model of survey response behavior and its implications for the analysis of different response effects will then be discussed in section 4. On the level of measurement theory, it is assumed that these paradata help to identify respondents answering in an automatic-spontaneous versus a deliberative-controlled mode of information processing. In the empirical analysis section (section 5), we estimate multiple group structural equation models to analyze measurement quality (validity

Contact information: Christoph Giehl, University of Kaiserslautern, Faculty of Social Sciences, Erwin-Schrödinger Str. 57, 67663 Kaiserslautern (E-mail: christoph.giehl@sowi.uni-kl.de) .

and reliability) of a four-item behavioral attitude scale (attitude towards health-conscious nutrition) with two positive and two negative items using CATI data with experimental variation of question order. Finally, we discuss our empirical findings (section 6).

2 Reversed Items in multiple-item scales

In survey research, it is a common and long-standing practice to operationalize latent constructs by a multiple-item approach with a mix of positively and negatively worded items (for an overview see Barnette, 2000). First of all, mixing the item directions helps to control for acquiescence bias on the level of research design by identifying respondents who do not answer substantially but just agree regardless of the question content. Assuming that the positively and negatively worded items are indicators of the same underlying latent construct, a valid measurement would be exhibited by consistent response behavior, i.e., agreeing to positively and disagreeing to negatively worded items (and vice versa). Moreover, answering several items on the same topic is a burden for respondents since they may perceive these questions as redundant and tedious. Therefore, mixing item directions is one of the most prominent methods to reduce perceived redundancy and control for acquiescence. Analysis of response behavior then allows survey researchers to identify valid and invalid response patterns in terms of acquiescence bias.

Unfortunately, mixing item directions may come with negative side effects with respect to measurement quality, which has been an ongoing discussion in the field of survey research for decades (Schriesheim & Hill, 1981). A lot of applications of principal component analysis and confirmatory factor analysis lead to artefactual factor structures where a simple one-factor structure does not fit the data (Marsh, 1986). In these models, negative items either generate their own substantial factor, i.e., a two-factor solution (Greenberger, Chen, Dmitrieva, & Farruggia, 2003; Spector, Van Katwyk, Brannick, & Chen, 1997) or they lead to the necessity of an additional latent method factor underlying the one-factor structure (e.g. Brown, 2003; Tomás, Oliver, Galiana, Sancho, & Lila, 2013; Weijters, Baumgartner, & Schillewaert, 2013; Wong, Rindfleisch, & Burroughs, 2003; Woods, 2006). A well-known example is Dunlap's New Environmental Paradigm Scale that displays a biased factor structure in which negative items do not load on same factor as positive items (Best & Mayerl, 2013; Dunlap, Van Liere, Mertig, & Jones, 2000). Brown (2003) found in his comparative work on mixed items that a one-factor solution containing an additional latent method factor is superior to a two-factor solution (which separates the positive and the negative dimensions), which in turn is superior to a one-factor solution without a latent method factor. In short, he concludes that factor analysis results are confounded or obscured by method bias (Brown, 2003, p. 1421) when not modeling these method-induced factorial structures for survey instruments with mixed items. In addition, researchers applying a mix of item directions have found lower scale reliability (Barnette, 2000), obtained different factor scores for items with the same content but reversed wording (Hartley, 2013), observed more positive opinions when items are worded negatively (Kamoen, Holleman, & van den Bergh, 2013), and experienced problems when adopting mixed items in cross-cultural studies (Wong et al., 2003). Thus, for the sake of parsimonious models, items with reversed direction are often recommended by researchers to simply be omitted (Hartley, 2013; Van Sonderen, Sanderman, & Coyne, 2013).

From a methodological perspective, these problems may arise from different sources. Firstly, survey researchers may face difficulties with the exact wording of reversed items (Hartley, 2013). Secondly, respondents may have problems with the reversed items, either being confused by having to reverse their thinking (disagreeing to a position would mean having to agree to a negatively worded item and vice versa) or just overlooking the different wording directions because of carelessness or individual tendency to acquiescence (Weijters et al., 2013; Woods, 2006). Thirdly,Weijters et al. (2013) discuss another possible mechanism they called confirmation bias, which describes the effect that the scale mean shifts towards the direction of the first item (but empirical results were inconclusive). Finally, problems may arise from non-attitudes, meaning that a lack of a "don't know" option could lead to satisficing strategies (J. A. Krosnick et al., 2002).

In the following sections, we provide some insights into problems of measurement quality of scales with reversed items via response latencies and analyze the main sources of these problems such as acquiescence.

3 Application of response latencies to analyze quality of survey data

The use of response latencies to measure cognitive processes has a long history in social psychology (Pachella, 1974) and was introduced into CATI research by Bassili and his colleagues (Bassili & Fletcher, 1991). Recently, response time itself has been in the focus of survey research papers (e.g Andersen & Mayerl, 2017; Couper & Kreuter, 2013; Mayerl, 2013; Olson & Smyth, 2015; Yan & Tourangeau, 2008). In a broader context, response latency measurement is a special type of non-reactive paradata (Couper & Kreuter, 2013). In general, many studies use response latencies as a proxy measure of chronic mental accessibility of political and social judgments (Cooke & Sheeran, 2013; Doll & Ajzen, 1992; Faas & Mayerl, 2010; Fazio, 1989; Marquis, 2014; Mayerl & Faas, 2018; Meyer & Schoen, 2014; Preißinger & Meyer, 2015; Ranger & Ortner, 2011; Urban & Mayerl, 2007). Another widespread application uses response latencies as a measure of the degree of elaboration, i.e., mode of information processing (Andersen & Mayerl, 2017; Baxter & Hinson, 2001; Carlston & Skowronski, 1986; Gibbons & Rammsayer, 1999; Hertel, Neuhof, Theuer, & Kerr, 2000; Mayerl, 2013; Schaffner & Roche, 2016; Sheppard & Teasdale, 2000; Shiv & Fedorikhin, 2002; Urban & Mayerl, 2007).

The measurement assumption of the latter application is that the higher the degree of elaboration, the longer the response time. This is due to the theoretical conceptualization that deliberate processing is a rational process of remembrance and consideration of individual information to generate a judgment (e.g Fazio, 1990b), which takes more time than processing an existing judgment or situational cues without reflection (Areni, Ferrell, & Wilcox, 1999; Brömer, 1999; Carlston & Skowronski, 1986; Doll & Ajzen, 1992; Gibbons & Rammsayer, 1999; Hertel & Bless, 2000; Hertel et al., 2000; Ruder, 2001; Smith, Fazio, & Cejka, 1996; Tormala & Petty, 2001). This assumption is supported by empirical evidence showing that an inconsistent information base causes longer reaction times (Bassili, 1996; Brömer, 1999; Klauer & Musch, 1999). Further studies have shown that the response latency increases with the amount of processed information (Bassili, 1996; Houlihan, Campbell, & Stelmack, 1994). Therefore, both the inconsistency and the amount of information affect judgments, only if the judgment is based on this information and is processed deliberately. Furthermore, since these properties of a judgment lead to longer reaction times, response latencies can be used as an indicator for the degree of elaboration.

Theory-driven applications have shown that response latency indeed acts as a moderator of stability and predictive power of social judgments where effects are stronger when latency is short (ee overview in Mayerl & Urban, 2008). Applications in the field of survey research have shown that response latencies act as a moderator of method effects like question order bias, showing that assimilation effect (also known as carry-over effect, see Tourangeau, Rasinski, Bradburn, & D'Andrade, 1989) is stronger in the case of fast times (Mayerl & Urban, 2008; Tourangeau, 1992) whereas contrast effect appears in the case of slow times (Mayerl & Urban, 2008; Stocké, 2002). Further studies showed that short response latencies are associated with the occurrence of acquiescence bias (Knowles & Condon, 1999; Mayerl, 2013), as well as, straight-lining behavior (Zhang & Conrad, 2013).

Response latency measurement can be divided into either interviewer-based active measurement via key pressing or completely automatic passive measurement where time stamps are recorded only (in this paper we focus on interviewer-based reaction time measurement). Interviewer-based measurement has the advantage that response times are more precise and interviewers are able to validate their time measurement. Since response latency is biased by characteristics of the measurement instrument, interviewers, survey situation, and individual characteristics of respondents then an appropriate data treatment of response latencies is necessary. This includes identification of invalid measurements (e.g. by outlier definition and timer validation of interviewers) and the control of individual baseline speed of the respondent (see Mayerl and Urban (2008) for a deeper discussion of different methods of measurement and data treatment).

It is important to note that while response effects like question order or acquiescence bias generate biased answers from respondents, there has been less discussion on whether such response effects indeed reduce or perhaps even enhance our measures of reliability or validity in survey research. This is especially true for measures of measurement quality, which rely on covariances between items (e.g. Cronbach's Alpha and factor loadings in confirmatory factor analysis), meaning that when response effects artificially enhance inter-item-covariances, measures of reliability and validity may be artificially enhanced too. Thus, we have to keep in mind which response effects affect measurement quality analysis in which direction. By establishing this, we expect assimilation effects of question order to enhance inter-item-covariances while contrast effects of question order would result in lower inter-item-covariances. Interestingly, the effect of acquiescence bias depends on the construction of the attitude scale itself. Hence, when the evaluative direction of question wording is the same for all items of the attitude scale, acquiescence is expected to strengthen inter-item-covariances. On the other hand, when introducing positive and negative question wording, acquiescence bias should result in lower inter-item-covariances.

4 Respondents' Behavior and Response effects in surveys

In social psychology and social cognition research, dual-mode models of different cognitive systems of information processing are very popular (**kahneman11**; see Chaiken & Trope, 1999; Mayerl, 2009; Smith & DeCoster, 2000). The most prominent dual-process models in attitude and survey research are the MODE model (Fazio, 1990b), the Elaboration Likelihood model (R. Petty & Cacioppo, 1986), the Heuristic-Systematic model (Chaiken, 1980), the Reflective-Impulsive model (Strack & Deutsch, 2004), and Krosnick's distinction of optimizing and satisficing routes (J. Krosnick, 1991). All these models assume two types of cognitive modes when people process information and generate judgments and attitudes. In survey research, these cognitive modes can be applied to two different kinds of responses (Mayerl, 2013): an automatic-spontaneous mode and a deliberative-controlled mode.

On the basis of dual-process models, the theoretical assumption underlying response time, as a measure of the degree of elaboration, is that a deliberative-controlled process is a rational decision process of retrieving arguments and beliefs and generating a judgment based on “raw” information. Therefore, this needs more time compared to automatic processing of chronic or temporarily highly accessible cues or heuristics (Areni et al., 1999; Brömer, 1999; Carlston & Skowronski, 1986; Doll & Ajzen, 1992; Gibbons & Rammsayer, 1999; Hertel & Bless, 2000; Hertel et al., 2000; Ruder, 2001; Smith et al., 1996; Tormala & Petty, 2001).

To gain a deeper understanding of the occurrence of response effects, it is crucial to distinguish between the types of response effects based on the mental effort, which is indicated by each biasing factor. In this sense, different response effects are to be expected depending on whether the cognitive response mode of information processing is automatic-spontaneous or deliberative-controlled. Spontaneous response behavior is characterized by cognitive processing of chronically accessible judgments (e.g. attitudes) or temporarily accessible information, which may stem from a particular situation or actual cognitive associations. Responses, which are based on temporarily accessible information, are highly susceptible to be biased by method effects that do not require a high degree of elaboration, e.g. question order assimilation effects (Tourangeau et al., 1989), acquiescence bias (Knowles & Condon, 1999), extreme responding or tendency to middle category (Yang, Harkness, Chin, & Villar, 2010). Furthermore, since short response latencies are a proxy measure of spontaneous cognitive processing, it comes as no surprise that these method effects have been found to occur when response time is short (see section 3).

On the other hand, method effects like comprehension problems, question interpretation and wording effects, as well as problems with the answering categories, come along with longer response latencies. This indicates that respondents had to think about an answer in a deliberative way by decoding the questions and possible answers. In addition, question order contrast effects (Schwarz & Strack, 1999) need a high degree of elaboration since the question content of the previous question has to be excluded or used as an anchor when answering a question. For this reason, contrast effects are found in the case of long response latencies (Stocké, 2002).

In summary, the occurrence of biasing response effects is associated with the response mode of information processing, which is activated during the answering process. Response mode processes, in turn, can be measured by response latencies. This is the theoretical and methodological background of our empirical analyses: - response latencies act as a moderator of measurement quality of an attitude scale with positive and negative items. Thus, we seek insights into different patterns and sources of measurement quality by distinguishing fast (spontaneous) versus slow (more elaborated) responses.

The following *three expectations* are derived from the discussion so far. In short, we state that different response effects occur for fast vs. slow responders which in turn lead to higher vs. lower measures of reliability and

1. For quick answers to attitudinal scales including reversed items, there is a higher chance of the occurrence of . . .
   - assimilation effects of question order leading to enhanced measures of reliability and validity, but also of the. . .
   - occurrence of acquiescence bias (as an indication of carelessness) leading to lower measures of reliability and validity.
2. Slow answers to attitudinal scales including reversed items imply . . .
   - respondents’ confusion and problems of understanding the reversed wording...
   - as well as possible contrast effects of question order. . .
   - leading to lower measures of reliability and validity.
3. Careful specification of measurement models with reversed items, including latent method factors, may help to control for response effects like item wording direction or acquiescence and thus enhance reliability and validity.

5 Empirical Analysis

5.1 Methods and data

Statistical analyses shown in this paper was derived from a German nation-wide CATI study with random sampling and experimental design. The total sample size was 2002 respondents with a net response rate of 28.4%. However, the analyses in this paper are focused on two experimental conditions with a total sample size of 379 (see below). Data were collected from February 2005 to April 2005 (Mayerl & Urban, 2008; Urban, 2005, for more details on this study).[1]

Using a 5-point rating scale, the attitude towards health-conscious nutrition is operationalized by four items, two with positive and two with negative wordings (see Appendix A1 for descriptive statistics of model variables): item a(+) “I personally think it is very good to eat in a health-conscious way.”, item b(-) “It is more of a disadvantage for me if I always eat in a health-conscious way.”, item c(+) “In everyday life I think it is especially good to eat in a health-conscious way.”, and item d(-) “Ultimately, it doesn’t make a lot of sense to eat exclusively in a health-conscious way.” For the ease of interpretation of the statistical results, all item values

[1] The data used in this paper will be made accessible by the principal investigator upon request, please contact the corresponding author for this purpose.

were recoded to 1: "not agree" to 5: "fully agree" towards health-conscious nutrition.

To control for effects of item order, the survey was conducted using an experimental design with a random assignment of respondents to one of two experimental conditions: random order of the four attitude items (group "random") versus fixed item order (group "fixed"). In the fixed question order condition, items appear in the order a-b-c-d, thus alternating between positive and negative items. The total sample size is 379 respondents, with 197 in the group "fixed" and 182 in the group "random".

Response latencies were actively measured in hundredths of seconds by interviewers for all items. In our analyses, we use the residual index (see Mayerl, Sellke, & Urban, 2005; Mayerl & Urban, 2008) to control for individual baseline speed. Individual baseline speed covers the "general mental speed that a person needs to answer questions, independent of the content of the question". (Mayerl, 2013, p. 4). In our study, baseline speed is operationalized as the mean time of three simple questions: vegetarian (yes/no), religious confession, and year of birth. These items are all highly, cognitively accessible. Thus, our measure controls for a basal physical-motoric baseline speed (see Mayerl et al., 2005 and Mayerl and Urban, 2008 for a comparison of different baseline speed measures). Taking individual baseline speed into account is necessary to control for a variety of methodological (e.g. interviewer speed), as well as individual (e.g. age, health) factors, which would otherwise disturb a proper interpretation of response latencies as a proxy of mental processes (see Fazio, 1990a; Mayerl & Urban, 2008, pp. 69). In addition, invalid time measurements are treated as missing values. Invalid time measurements are identified by two methods: interviewer validation and outlier treatment (outliers are defined as response times higher/lower than 2 times standard deviation of the mean). Overall, treating response times in this way enables the control of a wide range of biasing factors (see Mayerl & Urban, 2008).

For analyses on the construct level, the response latency of the attitude scale is aggregated as the mean response latency of the four items. In the next step, the median split was taken to differentiate "slow" vs. "fast" responders (Mayerl & Urban, 2008, see). This median split can be interpreted as a proxy of automatic-spontaneous (fast) versus deliberative-controlled (slow) response modes.

Figure 1 shows the structural equation models (SEM) we estimated with Mplus 7. First, we specify a simple model (Model 1), followed by a more complex model that includes method effects like item wording direction and acquiescence (Model 2). The latent method factor in model 2 covers negative wording of the two reversed items. This latent method factor is specified in line with CTC(M-1) models proposed by Eid, Schneider, and Schwenkmezger (1999) (CTCM=Correlated Trait Correlated Methods). Most importantly, in order to eliminate identification problems, this model specifies all possible method factors but omits one. Thus, in our case, we specify a latent method factor for all the negative items only. In order to investigate acquiescence as a source of reversed item bias, a manifest predictor variable "individual tendency to acquiescence (ACQU)" will be used as a control variable. In our study, individual tendency to acquiescence is measured by the number of items to which a respondent "fully agrees" out of exactly 100, in the 5-point rating scale items during the survey.[2] Effects of the individual tendency to acquiescence are expected to appear on the latent attitude construct and latent method factor, since attitude items were measured with a 5-point rating scale that asked how many respondents agree with the content. The effect of the acquiescence score on the method factor covers the expectation that the higher the tendency to acquiescence, the higher the tendency to agree to negative items. Moreover, since the negative items were recoded, we expect a negative effect of acquiescence on the latent method factor. On the other hand behavioral intention was measured as a percentage scale of the likelihood of specific nutrition behavior in the future. Consequently, we do not expect acquiescence effects here and specify this effect only for the sake of statistical control.

As all four items are approximately normally distributed in terms of skewness and multivariate kurtosis, the following models will be estimated with a ML estimator (see Byrne, 2010; Urban & Mayerl, 2014).[3] All multiple group models are specified without equivalence constraints of loadings and (co-)variances of the latent attitude construct in order to unveil measurement problems in each group.[4]

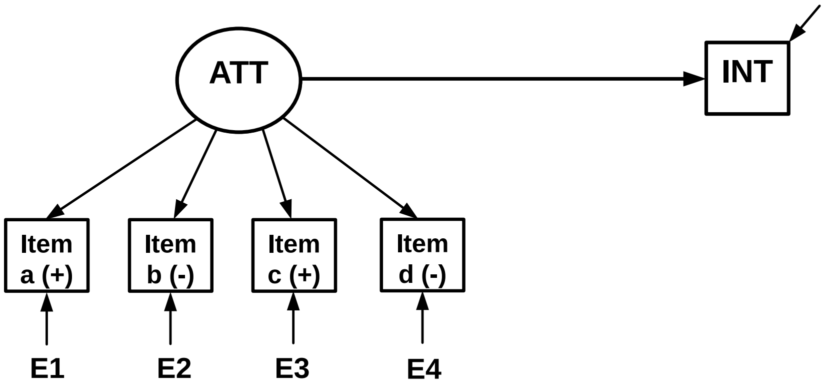
In estimating these structural equation models, we seek to fit an acceptable one-factor model for both fixed and random question order conditions in a first step (without response latency grouping), as well as looking more closely at fast and slow responders, in a second step. By estimating model 1 and model 2 with and without response latencies, SEM analyses help to evaluate whether measurement quality can be enhanced by a more complex model 2, as well as by introducing response latency groups.

[2]These items on multiple topics are about attitudes and perceived norms on environmental concern, religion, attitude towards surveys, health-conscious nutrition, and different psychological scales. Such a measure of acquiescence has already been tested and shown to be a valid measure of acquiescence (Mayerl, 2013; Watson, 1992).

[3]All models were re-run with a robust MLR estimator that shows no substantial differences in results when compared to the ML results presented in this paper.

[4]Due to free factor loadings in all groups, strength of causal effects of attitude on intention should not be compared between groups. But for purpose of our analysis of measurement quality, it is enough to know whether predictive validity of attitude factor is given or not.

a) Model 1



b) Model 2

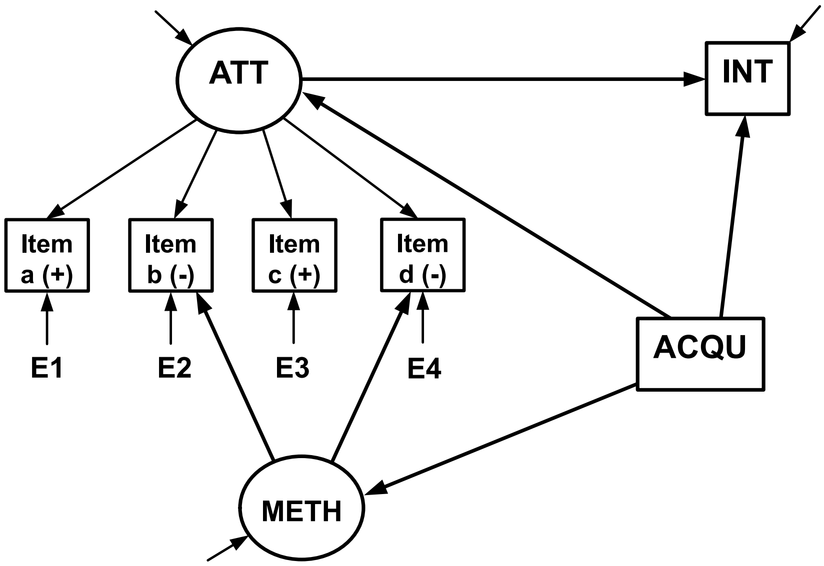


Figure 1. Latent Structural Equation Models without (M1) and with (M2) method effects

Note: Note.ATT=attitude; INT=behavioral Intention; MET=method factor; ACQU=individual acquiescence score; (+)=positive wording; (-)=negative wording

Since the paper aims to identify subgroups of high versus low measurement quality of the attitude scale, the following criteria of validity, reliability and model fit will be used to evaluate measurement quality (for more details on these criteria see Kline, 2016; Urban and Mayerl, 2014):

- Internal validity or convergent validity of each item:[5]
  - significant unstandardized factor loadings ($p<0.05$)
  - positive factor loadings (due to negative wording of some attitude items, item values were recoded to produce a rating scale that ranged from 1: "not agree" to 5: "fully agree" towards health-conscious nutrition)

[5]In our structural equation models, it is also possible to test for the absence of causal effects of attitude indicators on behavioral intention, i.e., to test for discriminant validity as a second part of construct validity.

- standardized factor loadings > 0.5

- Internal validity or convergent validity as composite measure:

  - AVE (Average Variance Extracted: mean of all squared standardized factor loadings) > 0.5 (less strict: > 0.4)

- Predictive validity as a special case of criterion validity:

  - replication of a well-known causal effect of attitude towards behavior scale on behavioral intention towards health-conscious nutrition (significant positive effect with substantial effect size). This effect is expected on the theoretical ground of the established and empirically tested Theory of Reasoned Action by Ajzen and Fishbein (see Ajzen & Fishbein, 1980)[6]

- Composite measure of reliability:

  - Jöreskog's ρ >0.7 (less strict: >0.6)
  - Cronbach's Alpha >0.7 (less strict: >0.6)

- Model Fit:

  - CFI > 0.95; RMSEA < 0.05 (90% Confidence interval upper bound <0.10); SRMR < 0.05; Chi^2 non-significant with p>0.05

5.2 Analysis of scale quality without latency differentiation

In order to evaluate the ability to distinguish between different subsamples of measurement quality using response latencies, it is crucial to first look at measurement quality without differentiation into subgroups of respondents. Due to experimental design of fixed vs. random item order, we are able to control for order effects which will – on aggregate – not appear in the condition of randomized order, but are likely to appear under the condition of fixed order. To be more precise, we expect that an assimilation context effect leads to higher correlations between two sequenced items, thus leading paradoxically to higher reliability and validity measures since these measures are based on correlations between the items. Table 1 reports item associations of the four attitude items in their order of appearance in the fixed question order group.

Table 1 shows that associations between subsequent items indeed are – at least in tendency – stronger in the case of a fixed item order. In two out of three cases, item relations are even non-significant in the case of random item order (p>0.05). This confirms our assumption that a question order assimilation effect is present on the level of simple item associations.

We now begin our main analyses estimating the simple SEM model 1 without any method factors.

Results in table 2 clearly show that attitude measurement in both groups is not acceptable in terms of convergent validity (AVE <0.4) and model fit (CFI <0.9 and RMSEA >0.1). The negative worded items do not measure the same underlying one-dimensional construct, indicated by standardized factor loadings far below 0.5. Reliability measures are lower than 0.7 but falls within an acceptable range. Surprisingly, scale quality is even worse in the group of random question order in contrast to fixed item order in terms of Jöreskog's and Cronbach's composite reliability measures. In addition, the factor loading of item b is even non-significant in the group of random item order (p>0.05). Predictive validity is given under both conditions since the attitude-intention effect is substantial and significant in both groups. In summation, the simple model 1 does not reach an acceptable level of scale quality in both groups, due to lack of construct validity and model fit.

Table 3 reports the results for the re-specified model 2 taking method effects into account.[7] In this instance, the measurement quality of our attitude scale is acceptable in terms of the criteria of validity and reliability (see section 5.1) under the condition of fixed item order but is still unacceptable in the case of random order. Moreover, the convergent validity is not given for random question order, since standardized factor loadings and AVE are still far too low to accept. Thus, it is the question order assimilation effect we found to be occurring in the fixed question order group (see table 2), which plays a crucial role as to whether scale quality is accepted or not. Hence, validity is remarkably higher when question order effects are likely to appear (fixed question order)[8] and validity is unacceptably low in the case of random order.

In terms of model fit, model 2 (table 3) is clearly superior to model 1 (table 2). This confirms the results of Brown (2003) discussed in section 2, which concludes that a method factor is effective in order to get an acceptable fit for latent

[6] Since specific nutrition intentions do *not* reflect a common latent construct of *subjective* health-conscious nutritional intention, behavioral intention is operationalized as an "*objective*" additive index according to recommendations of German health institutions in 2005 (likelihood scale 0...100%). The question wording is as follows: "What do you think, how likely is it that you ...
Item 1: ...will eat meat no more than 2 times per week in the next 4 weeks?
Item 2: ...will eat chocolate no more than 2 times per week in the next 4 weeks?
Item 3: ...will drink no more alcohol than 2 glasses of wine or beer per week?
Item 4: ...will eat fresh fruit and vegetable every day?"

[7] Unstandardized factor loadings of latent method factor are constrained to be equal over groups in order to identify multiple group models.

[8] The standardized factor loadings in the fixed order group are rather acceptable with only one item showing a factor loading slightly below 0.5 (item b: 0.45).

Table 1
Item relationships (linear regression models)

	Item order fixed[b] (n=197)			Item order random[c] (n=184)		
Items of attitude scale	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]
a (+) → Item b (-)	0.20*	0.09	0.15	0.07	0.09	0.05
b (-) → Item c (+)	0.12*	0.05	0.17	0.05	0.06	0.07
c (+) → Item d (-)	0.35*	0.10	0.26	0.29*	0.10	0.22

All item values recoded to 1: "not agree" ... 5: "fully agree" to health-conscious nutrition.
[a] standardized regression coefficient [b] fixed order of items: a(+), b(-), c(+), d(-)
[c] random: randomized item order
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 2
Std.Err.M results for Model 1 – without response latency grouping

		Item order fixed[b] (n=197)			Item order random[c] (n=182)		
Model estimates		Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]
ATT → INT		1.28**	0.21	0.50	0.86*	0.21	0.38
factor loadings ATT	Item a (+)	1.00	0.00	0.85	1.00	0.00	0.90
	Item b (-)	0.39**	0.12	0.25	0.13	0.11	0.10
	Item c (+)	0.83**	0.11	0.75	0.75**	0.14	0.70
	Item d (-)	0.52**	0.12	0.34	0.40**	0.12	0.28
Measures of model evaluation							
AVE			0.37			0.35	
Jöreskog's ρ			0.65			0.60	
Cronbach's Alpha			0.65			0.57	
Fit of multiple group model		CFI=0.816; RMSEA=0.187 (0.149–0.227); SRMR=0.099; Chi^2=76.166; df=10; p=0.000					

All item values recoded to 1: "not agree" ... 5: "fully agree" to health-conscious nutrition.
[a] standardized regression coefficient [b] fixed order of items: a(+), b(-), c(+), d(-)
[c] random: randomized item order
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

factors with mixed item.

Interestingly, the latent method factor for *negative wording* works fairly well under both conditions, and *tendency to acquiescence* shows medium to strong and highly significant effects on both the attitude construct and the method factor (p<0.01; table 3). The negative effect on the method factor can be seen as an indication that respondents were unaware of the negative wording and provides clear evidence that acquiescence is indeed a source of reversed item bias.

To sum up the results of model 2 (table 3), most survey researchers would probably accept the results of reliability and validity tests of our 4-item-attitude construct in the case of fixed question order but not in the case of random question order. In other words, response effects like question order effects lead to higher reliability and validity measures, which is not the intention of survey researchers when evaluating measurement quality. We address this result further in a later section.

5.3 Response Latency and scale quality

We now expand our model by differentiating between two groups of respondents with short versus long response latencies, which are used as a proxy measure of the use of two different response modes. As already discussed in sections 3 and 4, assimilation effects of question order are to be expected in the spontaneous mode group (indicated by higher inter-item-associations leading to higher factor loadings and measures of composite reliability under condition of fixed item order), while the deliberative mode group should not show any assimilation. Instead, the deliberative mode group could show tendencies towards contrast effects (indicated by lower inter-item-associations leading to lower factor loadings

Table 3
Std.Err.M results for Model 2 – without response latency grouping

		Item order fixed[b] (n=197)			Item order random[c] (n=182)		
Model estimates		Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]
ATT → INT		1.45**	0.29	0.53	0.88**	0.22	0.38
ACQU → INT		–0.12	0.15	–0.07	0.01	0.13	0.01
ATT →	Item a (+)	1.00	0.00	0.80	1.00	0.00	0.87
	Item b (-)	0.74**	0.18	0.45	0.22	0.12	0.16
	Item c (+)	0.95**	0.09	0.81	0.79**	0.12	0.72
	Item d (-)	0.89**	0.18	0.55	0.50**	0.13	0.35
METH →	Item b(-)	1.00	0.00	0.70	1.00	0.00	0.56
	Item d(-)	0.99**	0.23	0.71	0.99**	0.23	0.54
CQU → METH		–0.36**	0.09	–0.53	–0.19**	0.07	–0.34
ACQU → ATT		0.36**	0.04	0.62	0.26**	0.06	0.36
Measures of model evaluation							
AVE			0.45			0.36	
Jöreskog's ρ			0.75			0.63	
Cronbach's Alpha			0.65			0.56	
Fit of multiple group model		CFI=0.992; RMSEA=0.042 (0.000–0.092); SRMR=0.028; Chi²=14.626; df=11; p=0.200					

All item values recoded to 1: "not agree" ... 5: "fully agree" to health-conscious nutrition.
[a] standardized regression coefficient [b] fixed order of items: a(+), b(-), c(+), d(-)
[c] random: randomized item order
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

and measures of composite reliability under condition of random item order).

Indeed, table 4 shows that assimilation effects of item order appear in the automatic-spontaneous mode group only. As seen here, in two out of three cases, item association is non-significant when item order is random and item associations are higher for fixed versus random item order. In contrast, when response latencies are long, all associations are non-significant and tend to be lower in two out of three cases for fixed item order versus random item order. These results are in line with findings from the literature on response latency effects, showing that fast answers are "biased" by assimilation effects (see section 3). More specifically, the highest scores of measurement quality in terms of inter-item-associations are expected to appear in the group of "fixed order and short latencies".[9]

Now we turn our attention to the results for SEM model 1 and model 2 when distinguishing between our two latency groups within each experimental condition. From the results of table 5, we can clearly conclude that the specification of model 1 – without taking any method effects into account – is not acceptable in all four conditions. The negative items load weakly (or even negatively) and in some cases are non-significant on the latent attitude construct. Furthermore, the model fit is poor. Thus, model 1 cannot be accepted in either group.

Table 6 reports the results of the re-specified model 2, again taking response latency into account. Measures of composite reliability show that reliability is highest in the group of short latencies and fixed item order. In addition, validity is only acceptable under the condition of "fixed item order and short response latencies". This is particularly true for the factor loadings of the four items, which are not acceptable in the other three groups (loading of item b is extremely weak and non-significant in these three groups). In line with the findings reported in sections 3 and 4, acquiescence bias for negative worded items is strong and highly significant ($p \leq 0.01$) in case of short response latencies but not significant ($p > 0.05$) in the case of slow responders (see the effect ACQU → METH). This result shows the special

[9]Generally, in case of slow responders, table 3 shows a contrast effect of question order of item c(+) and item d(-), indicated by a stronger association between these items in the condition of random order. However, as the coefficients are non-significant in both groups (fixed vs. random item order), there is no clear empirical support for this effect.

Table 4
Item relationship and response latencies (linear regression models)

	Automatic-spontaneous mode (response latency short[d]) (n=189)						Deliberative-controlled mode (response latency long[e]) (n=190)					
	Item order fixed[b] (n=116)			Item order random[c] (n=73)			Item order fixed[b] (n=81)			Item order random[c] (n=109)		
Items of attitude scale	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]
a (+) → Item b (-)	0.35**	0.13	0.25	0.06	0.17	0.04	−0.03	0.14	−0.03	0.05	0.11	0.05
b (-) → Item c (+)	0.17**	0.06	0.25	0.09	0.08	0.12	0.01	0.09	0.01	0.01	0.08	0.01
c (+) → Item d (-)	0.57***	0.12	0.42	0.41**	0.15	0.31	−0.05	0.16	−0.04	0.15	0.13	0.11

All item values recoded to 1: "not agree" ... 5: "fully agree" to health-conscious nutrition. [a] standardized regression coefficient [b] fixed order of items: a(+), b(-), c(+), d(-) [c] random: randomized item order [d] short: response latency ≤ median [e] long: response latency > median
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 5
SEM results for Model 1 – with response latency grouping

		Response Latency Short[d]						Response Latency Long[e]					
		Item order fixed[b] (n=116)			Item order random[c] (n=73)			Item order fixed[b] (n=81)			Item order random[c] (n=109)		
Model estimates		Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]
ATT → INT		1.28**	0.21	0.55	1.12**	0.36	0.41	0.87	0.52	0.35	0.57*	0.26	0.30
ATT →	Item a (+)	1.00	0.00	0.92	1.00	0.00	0.86	1.00	0.00	0.91	1.00	0.00	0.98
	Item b (-)	0.51**	0.15	0.33	0.26	0.22	0.15	−0.03	0.16	−0.02	0.06	0.11	0.05
	Item c (+)	0.86**	0.09	0.83	1.04**	0.22	0.82	0.50	0.28	0.50	0.53**	0.20	0.57
	Item d (-)	0.73**	0.13	0.51	0.55*	0.22	0.33	−0.10	0.18	−0.07	0.25	0.14	0.21
Measures of model evaluation													
AVE			0.48			0.39			0.27			0.33	
Jöreskog's ρ			0.76			0.66			0.37			0.55	
Cronbach's Alpha			0.74			0.65			0.36			0.46	
Fit of multiple group model		CFI =0.836; RMSEA=0.179 (0.140–0.221); SRMR=0.099; Chi^2=80.982; df=20; p=0.000											

All item values recoded to 1: "not agree" ... 5: "fully agree" to health-conscious nutrition. [a] standardized regression coefficient
[b] fixed order of items: a(+), b(-), c(+), d(-) [c] random: randomized item order [d] short: response latency ≤ median [e] long: response latency > median
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

role of acquiescence, as a source of reversed item bias under the condition of spontaneous answering, and the need to control for these method effects. The overall model fit is now very good, confirming the need to specify a more complex factorial structure for mixed items.[10]

In summing up these results, we can conclude that a coarse overall analysis of measurement quality of all respondents at once would overlook strong problems of validity and reliability of attitude scales with positive and negative items, even when complex latent structural equation models are specified (see results presented in table 3, which could lead to the misleading conclusion that measurement quality is good for fixed question order). We have shown that response latency measurement helps to identify subgroups of high versus low measurement quality. However, the attitude scale with reversed items works fine under the condition of "fast responses *and* fixed item order" only. In addition, SEM analyses have shown that this is only true in the case when one explicitly specifies a method factor, which controls for question wording and acquiescence.

Interestingly, question order effects help to artificially enhance measurement quality in terms of our given measures of validity and reliability when used in empirical social research. This is an often, unrecognized side effect of response effects. In addition, specifying complex structural equation models with latent method factors helps to differentiate method effects like question wording and acquiescence from validity and reliability of the core latent construct (see improvement of model 2 in comparison to model 1). Analyses have shown that the specification of a latent method factor (model 2) is necessary and the global fit of a model without such a factor is unacceptable (model 1).[11]

Slow respondents appear to be confused by negative item wording, leading to low reliability and low validity measures. It is worth noting that measurement quality in terms of validity and reliability is lowest in the case of long latencies with fixed question order.

6 Conclusion

Measurement quality of survey-based attitude scales is an important topic in survey research. Since mixing negative and positive item wording is a common recommendation to control for acquiescence, we took a deeper look into such a mixed attitude scale by experimentally differentiating fixed vs. random item order, as well as, identifying subsamples of spontaneous vs. deliberate responders using response latencies.

In terms of measures of validity and reliability, we clearly find the following patterns: *Firstly*, the attitude construct with mixed items shows higher measures of data quality when item order is fixed, due to question order response effects. Thus, response effects may lead to higher scores of reliability and validity measures used in empirical social research. *Secondly*, an attitude scale including reversed items works fine for fast responders (i.e. spontaneous answers) while slow responders (i.e. deliberative answers) demonstrate very serious difficulties with the mixed wordings of the items. *Thirdly*, the highest levels of measurement quality are clearly reached under the condition of fast answers in combination with fixed item order. This finding implies that the "best" survey results are achieved when answers are spontaneously processed and question order effects are present. *Fourthly*, latent structural equation models showed evidence in favor of models including a latent method factor for "reversed or rather negative item wording". This confirms the need to control for response styles and other method effects covered by the latent method factor. *Fifthly*, the sources of this latent method factor are different for spontaneous versus deliberative modes of information processing. Hence, while the method factor is strongly dependent on acquiescence bias when respondents answer questions in a spontaneous way, this is not the case for deliberative answering processes.

In our search for an answer, as to whether mixing item directions should be recommended or not, we could come up with no clear suggestion. On the level of research design, scales with reversed item wordings may allow researchers to identify acquiescence bias but at the same time, far more complex structural equation models with latent method factors must be estimated in order to achieve an acceptable operationalization of the latent construct. Our analyses have shown that paradata, like response latency measurements help to identify measurement problems in surveys, which would have been overlooked with traditional survey analysis strategies. Thus, response latencies have the strong potential to help survey researchers gain a deeper understanding of respondents' behavior, as well as, measurement quality.

Acknowledgements

The authors are indebted to Prof. Dr. Dieter Urban (University of Stuttgart, Germany) for providing the data conducted in the research project "ARIS – Response Latency Measurement in Surveys. Analyzing the cognitive basis of attitudes and information processing", granted by the

[10]In addition, modification indices do not indicate any additional significant path in the model in each group, i.e. there is no link missing between attitude indicators and behavioral intention. This can be interpreted as acceptable discriminance between operationalizations of attitude and intention.

[11]In addition, a bi-dimensional factor solution was estimated: one latent factor with positive and one with negative items. While the model fit is as good as the fit of our method-factor model (model 2), standardized factor loadings of negative items are above 1.0 in two groups and the factor with negative items does not have a substantial and significant effect on behavioral intention in all four groups. Thus, in line with Brown (2003), we conclude that a latent method factor model is superior to a two-factor model.

Table 6
SEM results for Model 2 – with response latency grouping

		Response Latency Short[d]						Response Latency Long[e]					
		Item order fixed[b] (n=116)			Item order random[c] (n=73)			Item order fixed[b] (n=81)			Item order random[c] (n=109)		
Model estimates		Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]	Coef.	Std.Err.	Coef.[a]
ATT → INT		1.46^{**}	0.27	0.60	1.07^{**}	0.37	0.38	1.13	0.90	0.33	0.71^{**}	0.27	0.34
ACQU → INT		–0.19	0.16	–0.12	0.10	0.17	0.07	0.09	0.35	0.06	–0.12	0.19	–0.06
ATT →	Item a (+)	1.00	0.00	0.88	1.00	0.00	0.84	1.00	0.00	0.66	1.00	0.00	0.91
	Item b (-)	0.80^{**}	0.18	0.50	0.35	0.23	0.20	0.14	0.39	0.08	0.13	0.13	0.11
	Item c (+)	0.94^{**}	0.08	0.87	1.08^{**}	0.21	0.84	0.95^{**}	0.22	0.68	0.62^{**}	0.15	0.62
	Item d (-)	1.09^{**}	0.17	0.74	0.70^{**}	0.23	0.41	0.23	0.44	0.12	0.36^{*}	0.16	0.28
METH →	Item b (-)	1.00	0.00	0.61	1.00	0.00	0.62	1.00	0.00	0.64	1.00	0.00	0.40
	Item d (-)	1.15^{**}	0.27	0.76	1.15^{**}	0.27	0.74	1.15^{**}	0.27	0.68	1.15^{**}	0.27	0.43
ACQU → METH		-0.37^{**}	0.10	–0.57	-0.22^{*}	0.09	–0.39	–0.14	0.15	–0.26	–0.16	0.09	–0.37
ACQU → ATT		0.37^{**}	0.06	0.57	0.16^{*}	0.07	0.30	0.34^{**}	0.07	0.74	0.33^{**}	0.09	0.37
Measures of model evaluation													
AVE			0.58			0.41			0.23			0.33	
Jöreskog's ρ			0.84			0.69			0.44			0.58	
Cronbach's Alpha			0.74			0.63			0.36			0.46	
Fit of multiple group model		CFI=1.00; RMSEA=0.000 (0.000–0.083); SRMR=0.034; Chi^2 =22.591; df=23; p=0.485											

All item values recoded to 1: "not agree" ... 5: "fully agree" to health-conscious nutrition. [a] standardized regression coefficient
[b] fixed order of items: a(+), b(-), c(+), d(-) [c] random: randomized item order [d] short: response latency ≤ median [e] long: response latency > median
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

DFG ("Deutsche Forschungsgemeinschaft") from 1/1/2014 to 31/12/2005.

References

Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. London: Prentice-Hall.

Andersen, H. & Mayerl, J. (2017). Social desirability and undesirability effects on survey response latencies. *Bulletin of Sociological Methodology*, *135*(1), 68–89.

Areni, C. S., Ferrell, M. E., & Wilcox, J. B. (1999). The effects of need for cognition and topic importance on the latency and extremity of responses to attitudinal inquiries. *Advances in Consumer Research*, *26*, 63–68.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361–370.

Bassili, J. N. (1996). The how and why of response latency measurement in telephone surveys. In N. Schwarz & S. Sudman (Eds.), *Answering Questions. Methodology for determining cognitive and communicative process in survey research*. San Francisco: Jossey-Bass.

Bassili, J. N. & Fletcher, J. F. (1991). Response-time measurement in survey research a method for CATI and a new look at nonattitudes. *Public Opinion Quarterly*, *55*(3), 331–346.

Baxter, B. W. & Hinson, R. E. (2001). Is smoking automatic? Demands of smoking behavior on attentional resources. *Journal of Abnormal Psychology*, *110*(1), 59–66.

Best, H. & Mayerl, J. (2013). Values, beliefs, attitudes: An empirical study on the structure of environmental concern and recycling participation. *Social Science Quarterly*, *94*(3), 691–714.

Brömer, P. (1999). *Informationsverarbeitung bei ambivalenten Einstellungen*. Regensburg: Roderer.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy*, *41*(12), 1411–1426.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, London: Routledge.

Carlston, D. E. & Skowronski, J. J. (1986). Trait memory and behavior memory: The effects of alternative pathways on impression judgment response times. *Journal of Personality and Social Psychology*, *50*(1), 5–13.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752–766.

Chaiken, S. & Trope, Y. (1999). *Dual-process theories in social psychology*. New York, London: Guilford Press.

Cooke, R. & Sheeran, P. (2013). Properties of intention: Component structure and consequences for behavior, information processing, and resistance. *Journal of Applied Social Psychology*, *43*(4), 749–760.

Couper, M. P. & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 271–286.

Doll, J. & Ajzen, I. (1992). Accessibility and stability of predictors in the theory of planned behavior. *Journal of Personality and Social Psychology*, *63*(5), 754–756.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). Measuring endorsement of the new ecological paradigm: A revised NEP scale. *Journal of Social Issues*, *56*(3), 425–442.

Eid, M., Schneider, C., & Schwenkmezger, P. (1999). Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality*, *13*(4), 283–306.

Faas, T. & Mayerl, J. (2010). Michigan Reloaded. Antwortlatenzzeiten als Moderatorvariablen in Modellen des Wahlverhaltens. In T. Faas, K. Arzheimer, & S. Roßteutscher (Eds.), *Information – Wahrnehmung – Emotion: Politische Psychologie in der Wahl- und Einstellungsforschung* (pp. 259–276). Wiesbaden: VS Verlag.

Fazio, R. H. (1989). On the power and functionality of attitudes: The role of attitude. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude, structure and function* (pp. 153–179). Erlbaum.

Fazio, R. H. (1990a). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Review of personality and social psychology* (Vol. 11, pp. 74–97). Review of Personality and Social Psychology. Newbury Park, CA: Sage Publications.

Fazio, R. H. (1990b). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, *23*, 75–109.

Gibbons, H. & Rammsayer, T. (1999). Auswirkungen der Vertrautheit mit einer Reizdimension auf Entscheidungsprozesse: Der modulierende Einfluss kontrollierter vs. automatischer Informationsverarbeitung. In I. Wachsmuth & B. Jung (Eds.), *KogWis99, Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft* (pp. 159–164). Bielefeld, St.Augustin.

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, *35*(6), 1241–1254.

Hartley, J. (2013). Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology*, *13*, 83–86.

Hertel, G. & Bless, H. (2000). “On-line” und erinnerungsgestützte Urteilsbildung: Auslösefaktoren und empirische Unterscheidungsmöglichkeiten. *Psychologische Rundschau*, *51*(1), 19–28.

Hertel, G., Neuhof, J., Theuer, T., & Kerr, N. L. (2000). Mood effects on cooperation in small groups: Does positive mood simply lead to more cooperation? *Cognition & Emotion*, *14*(4), 441–472.

Houlihan, M., Campbell, K., & Stelmack, R. M. (1994). Reaction time and movement time as measures of stimulus evaluation and response processes. *Intelligence*, *18*(3), 289–307.

Kamoen, N., Holleman, B., & van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, *7*(3), 181–189.

Klauer, K. C. & Musch, J. (1999). Eine Normierung unterschiedlicher Aspekte der evaluativen Bewertung von 92 Substantiven. *Zeitschrift für Sozialpsychologie*, *30*(1), 1–11.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Knowles, E. S. & Condon, C. A. (1999). Why people say “yes”: A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*(2), 379–386.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., … Conaway, M. (2002). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, *66*(3), 371–403.

Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213–236.

Marquis, L. (2014). The psychology of quick and slow answers: Issue importance in the 2011 Swiss Parliamentary Elections. *Swiss Political Science Review*, *20*(4), 697–726.

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*(1), 37.

Mayerl, J. (2009). *Kognitive Grundlagen sozialen Verhaltens-Framing, Einstellungen und Rationalität*. Wiesbaden: VS Verlag.

Mayerl, J. (2013). Response latency measurement in surveys. Detecting strong attitudes and response effects. *Survey Methods: Insights from the Field (SMIF)*. Retrieved from website:%20http://surveyinsights.org/?p=1063.

Mayerl, J. & Faas, T. (2018). Campaign dynamics of cognitive accessibility of political judgments: Measuring the impact of campaigns and campaign events using response latencies in two German rolling cross section studies. *Quality & Quantity*, *52*(4), 1575–1592.

Mayerl, J., Sellke, P., & Urban, D. (2005). *Analyzing cognitive processes in CATI-surveys with response latencies: an empirical evaluation of the consequences of using different baseline speed measures*. Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart, SISS No. 2/2005.

Mayerl, J. & Urban, D. (2008). *Antwortreaktionszeiten in Survey-Analysen*. Wiesbaden: VS Verlag.

Meyer, M. & Schoen, H. (2014). Response Latencies and Attitude-Behavior Consistency in a Direct Democratic Setting: Evidence from a Subnational Referendum in Germany. *Political Psychology*, *35*(3), 431–440.

Olson, K. & Smyth, J. D. (2015). The effect of CATI questions, respondents, and interviewers on response time. Sociology Department, Faculty Publications. Paper 268. Retrieved from http://digitalcommons.unl.edu/sociologyfacpub/268.

Pachella, R. G. (1974). The interpretation of reaction time in human information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale: Erlbaum.

Petty, R. & Cacioppo, J. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 124–205.

Preißinger, M. & Meyer, M. (2015). Einstellungszugänglichkeit im Laufe von Wahlkämpfen: Aktivierungseffekte im Kontext der Bundestagswahlen 2005, 2009 und 2013. In T. Faas, C. Frank, & H. Schoen (Eds.), *Politische Psychologie* (pp. 444–469). PVS Sonderheft 50. Baden-Baden: Nomos.

Ranger, J. & Ortner, T. M. (2011). Assessing personality traits through response latencies using item response theory. *Educational and Psychological Measurement*, *71*(2), 389–406.

Ruder, M. (2001). Die Nutzung der Verfügbarkeitsheuristik-eine Frage der Stimmung? In *Studienreihe Psychologische Forschungsergebnisse, Band 79*. Hamburg: Kovac.

Schaffner, B. F. & Roche, C. (2016). Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opinion Quarterly*. First published online: December 8, 2016. doi:10.1093/poq/nfw043

Schriesheim, C. A. & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, *41*(4), 1101–1114.

Schwarz, N. & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahnemann, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Russell-Sage.

Sheppard, L. C. & Teasdale, J. D. (2000). Dysfunctional thinking in major depressive disorder: A deficit in metacognitive monitoring? *Journal of Abnormal Psychology*, *109*(4), 768–776.

Shiv, B. & Fedorikhin, A. (2002). Spontaneous versus controlled influences of stimulus-based affect on choice behavior. *Organizational Behavior and Human Decision Processes*, *87*(2), 342–370.

Smith, E. R. & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108–131.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, *71*(5), 888–898.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659–677.

Stocké, V. (2002). Die Vorhersage von Fragenreihenfolgeeffekten durch Antwortlatenzen: Eine Validierungsstudie. *ZUMA Nachrichten*, *26*(50), 26–53.

Strack, F. & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247.

Tomás, J. M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(2), 299–313.

Tormala, Z. L. & Petty, R. E. (2001). On-line versus memory-based processing: the role of "need to evaluate" in person perception. *Personality and Social Psychology Bulletin*, *27*(12), 1599–1612.

Tourangeau, R. (1992). Context effects on responses to attitude questions: Attitudes as memory structures. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 35–47). New York: Springer.

Tourangeau, R., Rasinski, K. A., Bradburn, N., & D'Andrade, R. (1989). Carryover effects in attitude surveys. *Public Opinion Quarterly*, *53*(4), 495–524.

Urban, D. (2005). Antwort-Reaktionszeiten in Surveys (ARIS) [Response Latency Measurement in Surveys]. Funded by the Deutsche Forschungsgemeinschaft. University of Stuttgart, Germany. Contact address: dieter.urban@sowi.uni-stuttgart.de.

Urban, D. & Mayerl, J. (2007). Antwortlatenzzeiten in der survey-basierten Verhaltensforschung. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *59*(4), 692–713.

Urban, D. & Mayerl, J. (2014). *Strukturgleichungsmodellierung: Ein Ratgeber für die Praxis*. Wiesbaden: Springer VS Verlag.

Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS ONE*, *8*(7), e68967. doi:10.1371/journal.pone.0068967

Watson, D. (1992). Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness. *Sociological Methods & Research*, *21*(1), 52–88.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*(3), 320–334.

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research*, *30*(1), 72–91.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186.

Yan, T. & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*(1), 51–68.

Yang, Y., Harkness, J. A., Chin, T.-Y., & Villar, A. (2010). Response styles and culture. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, . . . T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 203–223). Hoboken: Wiley.

Zhang, C. & Conrad, F. (2013). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, *8*(2), 127–135.

Appendix

(Appendix tables on next page)

Table A1
Descriptive statistics of model variables

Subgroup	Attitude Item a (+) (1-5)[a]	Attitude Item b (-) (1-5)[a]	Attitude Item c (+) (1-5)[a]	Attitude Item d (-) (1-5)[a]	Intention (INT) 0-100%	Acquiescence (ACQU) 0-100
fixed - slow						
N	81	81	81	81	81	81
Mean	2.22	2.58	2.41	2.47	55.11	20.75
Stddev	1.01	1.21	0.92	1.30	22.94	14.22
Min	1.00	1.00	1.00	1.00	10.00	1.00
Max	5.00	5.00	5.00	5.00	100.00	71.00
Skew	0.43	0.26	0.08	0.51	−0.05	1.20
Kurtosis	−0.26	−0.91	−0.33	−0.76	−0.84	1.63
Normalized Mardia coeff. (multivariate kurtosis): 1.943						
fixed - fast						
N	116	116	116	116	116	116
Mean	2.02	2.48	2.08	2.16	59.66	21.27
Stddev	1.03	1.46	0.97	1.34	21.96	13.94
Min	1.00	1.00	1.00	1.00	7.50	1.00
Max	5.00	5.00	5.00	5.00	100.00	73.00
Skew	0.89	0.43	0.60	0.85	−0.04	1.15
Kurtosis	0.29	−1.28	−0.09	−0.54	−0.28	1.68
Normalized Mardia coeff. (multivariate kurtosis): 2.274						
random - slow						
N	109	109	109	109	109	109
Mean	2.40	2.19	2.33	2.39	52.25	16.80
Stddev	1.15	1.25	1.04	1.35	21.52	11.55
Min	1.00	1.00	1.00	1.00	0.00	0.00
Max	5.00	5.00	5.00	5.00	100.00	66.00
Skew	0.65	0.84	0.62	0.58	−0.21	1.13
Kurtosis	−0.14	−0.20	0.19	−0.85	−0.38	2.05
Normalized Mardia coeff. (multivariate kurtosis): 6.963						
random - fast						
N	73	73	73	73	73	73
Mean	1.99	1.99	1.97	1.88	59.76	22.03
Stddev	0.96	1.39	1.04	1.39	22.33	15.57
Min	1.00	1.00	1.00	1.00	7.50	5.00
Max	5.00	5.00	5.00	5.00	100.00	85.00
Skew	0.79	1.21	0.82	1.52	−0.17	1.57
Kurtosis	0.18	0.06	−0.18	0.88	−0.58	3.19
Normalized Mardia coeff. (multivariate kurtosis): 3.942						

[a] all item values recoded to 1: “not agree” ... 5: “fully agree” to health-conscious nutrition; fixed: fixed order of items: a(+), b(-), c(+), d(-); random: randomized item order; Normalized Mardia coefficient estimated with AMOS 24