International Journal of Public Opinion Research Vol. 22 No. 3

doi:10.1093/ijpor/edq001 Advance Access publication 23 July 2010

VARIATIONS IN RESPONSE STYLE BEHAVIOR BY RESPONSE SCALE FORMAT IN ATTITUDE RESEARCH

Natalia D. Kieruj and Guy Moors

Tilburg University, The Netherlands

ABSTRACT

Studies concerning the impact of the length of response scales on the measurement of attitudes have primarily focused on the method bias associated with question format. At the same time another line of research has focused on the issue of response styles that affect how respondents answer to attitude questions. So far, research has paid less attention to the issue of whether the length of the response scales is related to response styles. In this study, we explore if differences in length of the response scale (*i.e.*, method factor) have differential effects in evoking extreme and midpoint response style behavior (*i.e.*, style factor). Our hypotheses read as follows. As the number of response categories increases, we expect subjects to be more likely to exert extreme response style. Furthermore, we expect subjects to be more likely to adopt a midpoint response style when they are offered a middle response category. To investigate these hypotheses we developed a split ballot experiment in which the number of response categories is manipulated from 5 to 11 categories. Data are collected by a random sample, large-scale web survey which allows for random assignment to the experimental conditions. The results show clear evidence of extreme response style and moderate evidence of midpoint response style. Extreme response style is not affected by the length of response scales, whereas the exertion of midpoint response style only popped up in the longer scale versions.

It is well known in attitude measurement that question format can greatly influence the way subjects respond to attitude questions (Krosnick & Berent, 1993; Schwarz, 1999; Tourangeau & Smith, 1996; Van Herk, Poortinga, &

Corresponding author: Natalia D. Kieruj, Tilburg University, FSW-MTO, Room P. 1.103, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands; e-mail: n.d.kieruj@uvt.nl

Verhallen, 2004). Seemingly minor details of response scales can lead to systematic method error, which in turn obscures the measurement of the attitude of interest. It is equally well known that response styles such as extreme response style (ERS) and midpoint response style (MRS) can alter results in several nontrivial ways when measuring attitudes (Arce-Ferrer & Ketterer, 2003; Baumgartner & Steenkamp, 2001). ERS is the tendency of respondents to choose the extreme endpoints of a rating scale (Hurley, 1998), whereas MRS is the tendency to make disproportionate use of the middle response category (Weijters, 2006).

In this study we examine whether the length of a response scale, *i.e.*, the method used, is related to the occurrence of response style behavior. The effect of varying the number of response categories has been researched in the past, however, with little reference to the issue of response style behavior. For example, Miller (1956) linked the typical use of 7-point rating scales to the amount of information people are able to maintain in the span of immediate memory (which happens to be 7). Alwin and Krosnick (1991) found that increasing the number of answering categories led to higher reliabilities. In line with this, Alwin (1997) found that questions with more categories are more reliable and more valid. Similarly, Scherpenzeel and Saris (1997) have researched the expected level of validity and reliability of any given scale, as a function of the number of answering categories, by means of Multitrait-Multimethod models (MTMM).

The novelty of our research is that it links variations in scale format (*i.e.*, method factor) to response style behavior (*i.e.*, style factor). To the best of our knowledge there is little research on this topic. Scrolling through the literature we found quite some references about response format and its relationship to measurement issues, however, with little reference to response styles. Similarly, the literature on response styles rarely discusses the role of variations in response scale formats. Hence, this research aims at bridging the two lines of research.

The article is organized as follows. We first review the literature regarding the number of response categories and the two types of response style. Secondly, we describe the split ballot experimental design that is used to explore the relationship between response formats and response behaviors and present the latent class method used for analyzing response bias. Results and conclusions are reported afterward.

LITERATURE REVIEW

Distinguishing Method and Style

In the literature review we bring together some of the significant propositions and findings from within each line of research. First, we focus on the issue of

the length of response scales. Second, we discuss measurement issues related to ERS and MRS. Before doing this, however, we want to elaborate on a distinction between method and response style effects. Method effects are defined as systematic variance that is attributable to the measurement method rather than to the constructs the measures represent (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Response styles on the other hand, can be defined as a person's tendency to respond systematically to questionnaire items on some basis other than what the items were specifically designed to measure (Paulhus, 1991). Following from these definitions, method effects are attributable to the method itself, whereas response styles appear to be person related features. Although conceptually distinct, some amount of overlap amongst method effects and response styles occurs when the tendency to use response styles is attributable to the measurement method. For example, if a questionnaire openly inquires after a sore subject, this might lead some respondents to use a social desirability response style or it might enhance MRS. However, response styles can also occur independent of certain properties of the method, since they can also reside within the respondent. In other words, some respondents might just be more inclined to use a particular response style than others. Nevertheless, we think it is important to make the conceptual distinction between measurement issues related to method factors and issues related to response style factors.

The Length of Response Scales and Middle Answers

Varying the number of response options naturally leads to two potentially important variations in rating scales, *i.e.*, variations in length as well as variations in the presence or absence of a middle response option. Both these aspects of varying the number of response options will be examined in this study and in this section we will discuss some of the research that has been done on each of these scale aspects.

Likert scales (1932) are still very popular in public opinion research and since Likert originally proposed five response categories many large-scale surveys have adopted this format. However, the optimal amount of response options is not exactly settled and probably depends on numerous variables like question content and respondent factors. Also, the optimal amount of response options might differ depending on which measure is used. For example, the optimal amount in terms of response style behavior may differ from the optimal amount in terms of reliability.

Reliability is the measure on which is focused in most studies concerning the number of answering categories. As one of the first to examine the reliability issue, Symonds (1924) suggested that a 7-point rating scale is the best option. By now, it seems to be the consensus that reliability increases as the number of answering categories increases (Muñiz, García-Cueto,

& Lozano, 2005; Preston & Colman, 2000; Weng, 2004). Several researchers found that rating scales consisting of five categories begin to produce satisfactory reliability values (Preston & Colman, 2000; Weng, 2004). Adding categories to 5-point rating scales increases their reliability until a certain point is reached after which the advance comes to a halt. A considerable amount of studies show that this point is reached when 7-point scales are used (Alwin, 1992; Cicchetti, Shoinralter, & Tyrer, 1985; Preston & Colman, 2000). After staying constant for a while when additional response options are added, reliability tends to decrease again. For example, Preston and Colman (2000) found that this was the case when 11 response categories were used. Taken together, these studies seem to indicate that scales with five to seven answering categories are preferable, something that has been advocated by Krosnick and Fabrigar (1997) as well.

There have also been a few studies in which validity is specified as a function of the number of response options. Most of them show that validity of the test increases as the number of response options increases (Muñiz, García-Cueto, & Lozano, 2005; Preston & Colman, 2000; Thomas, Uldall, & Krosnick, 2004). Hence, it becomes apparent that some research has been done concerning the length of response scales. However, as said before, up until now none of these studies linked the length of response scales to response styles.

Besides length of response scales we also focus on the presence of a middle answer in these scales. When it comes to middle answers, the most important issue is whether or not a neutral response option offers an easy out for subjects who do not want to choose sides or if such an option is necessary to measure attitudes accurately (O'Muircheartaigh, Krosnick & Helic, 2000). Scholars have reached no clear consensus when it comes to this question. Many researchers found that including a neutral middle response option to a rating scale attracts subjects disproportionally to this category (Kalton, Roberts & Holt, 1980; Raaijmakers, Van Hoof, 't Hart, Verbogt, & Vollebergh, 2000; Si & Cullen, 1998). The results that O'Muircheartaigh, Krosnick and Helic (2000) obtained in their research, however, were in contrast with this finding. They reasoned that if subjects need the middle option to express their opinions optimally, they would only check this option if it was accurate and therefore select a random response option if the middle option were omitted. They found that offering the middle option led to higher reliabilities and less random method error compared to omitting this option, which led to their conclusion that the middle option is in fact crucial to measure opinions accurately. Also, Saris (1988) stated that midpoints may serve as an anchor to respondents which could add to data quality and Borgers, Hox and Sikkel (2004) found that omitting middle response options led to a decrease in reliability. Whether response style behavior is related to the presence or absence of a middle response option is not yet specified.

ERS AND MRS

Current research on response styles in attitude research has primarily focused on acquiescence (or agreement bias) and extreme response style. By its very nature the question of response style behavior is exploratory, *i.e.*, 'does it occur in a given dataset?' Exploring the data used in this research provided evidence of ERS and partial evidence of MRS. The former was not a huge surprise given that part of the questionnaire of this research was founded on previous research that discussed ERS (Moors, 2008). The occurrence of MRS, on the other hand, was less anticipated. In this section we will discuss both response styles in some detail.

It has been argued that ERS can lead to serious contamination of the observed scores in a dataset (Baumgartner & Steenkamp, 2001). More specifically, ERS skews score frequency distributions toward the extreme endpoints of a rating scale. This leads to increased variance, which in turn reduces correlation coefficients (Clarke, 2001; Hui & Triandis, 1989). Baumgartner and Steenkamp (2001) indicated that ERS led to stylistic variance in their dataset and that this led to bias in correlations between scales. Moors (2003) found that ERS influences the effect of covariates on attitudes in latent class factor structural equation models. In another study Arce-Ferrer and Ketterer (2003) showed that the factor structure obtained when using a sample with respondents high in ERS substantially departed from the structure obtained when using a sample with respondents low in ERS. The most critical issue, however, is that ERS can negatively influence the validity of the measurement of attitudes. For example, Arce-Ferrer and Ketterer (2003) demonstrated that ERS seriously distorts construct validity. Concerns about the validity of attitude measurement when ERS is involved are also expressed in the aforementioned references of Baumgarter and Steenkamp (2001) and Moors (2003).

The main topic of this research is on whether and when ERS occurs given the response format that is used. Hence, research on what causes the use of ERS is less relevant. Nevertheless, some findings on this issue are worth reporting. For example, many researchers investigating ERS link this response style to culture (Dolnicar & Grün, 2007; Hui & Triandis, 1989; Johnson, Kulesa, Cho, & Shavitt, 2005; Marín, Gamba, & Marín, 1992; Van Herk, Poortinga, & Verhallen, 2004). Others (Austin, Deary, & Egan, 2006) have linked the exertion of ERS to certain psychological characteristics of subjects such as extraversion or conscientiousness. Closer to the topic of this research, Krosnick (1991) has pointed out that measurement factors can influence the use of ERS as well.

Until now, there have been few studies explicitly investigating MRS. This might be the case because MRS is sometimes considered to be the counterpart of ERS (Hurley, 1998). However, although MRS and ERS

seem to be negatively correlated in many situations, this is not always the case (Stening & Everett, 1984; Weijters, 2006). Studies that did focus on MRS deal mainly with cultural differences in the exertion of MRS (Hamid, Lai, & Cheung, 2001; Mandal, Ida, Harizuka, & Upadhaya, 1999; Si & Cullen, 1998).

We hypothesize that extreme response style will become more pronounced as the number of answering categories increases. We base our expectation in part on Krosnick's concept of satisficing (1991). Krosnick's basic argument is that task difficulty is one of the factors that influence a respondent's tendency to satisfice. The latter implies that respondents, who are not willing to expand the necessary effort and time to form optimal answers to attitude questions, might choose to use heuristic shortcuts to formulate answers that satisfy them enough. Increasing the length of response styles might increase task difficulty in our study, therefore leading respondents to satisfice in the form of response style behavior. Complementary to the idea of satisficing is the finding by Weathers, Sharma, and Niedrich (2005) that as the number of scale points increases, the likelihood of only choosing a limited number of these response categories in a set of questions also increases. This suggests that when the actual rating scale is stretched too widely, respondents simplify their answering process by choosing certain anchor points of the rating scale and only use these scale points. Both the concept of 'satisficing' as well as 'anchor point search' lead to the expectation that response style behavior is evoked by the method used (scale length in the case of this research). If, however, response styles are much more a kind of personality trait—as has been suggested by Billiet and Davidov (2008) for instance—rather than the consequence of task difficulty, it remains to be seen how such a personality trait affects responding to differences in response scale length.

There are two possible solutions for dealing with response styles. One of them is to prevent subjects from exerting them. To do this it should be clear what it is exactly that provokes response styles. For example, one could try to determine the optimal amount of response categories that leads to the least possible response style behavior. However, as said before, this optimal amount is probably different for every situation. Furthermore, there are many other possible causes for response style behavior. Eliminating all of them would be very difficult if not impossible—especially when response style is part of a personality trait. Therefore, in this study, we opt for a second way of dealing with response styles, *i.e.*, by trying to detect them in the dataset at hand and control for their effect while measuring attitudes of interest. This is done by isolating a response style factor from the 'true' content of the attitude scales so that response styles distort the results as little as possible. The goal therefore, is to determine when response style behavior is easiest to detect, so that it becomes easier to correct for this kind of bias.

METHOD AND DATA

Participants

Our split ballot sample experiment was implemented in the MESS project (Advanced Multi-Disciplinary Facility for Measurement and Experimentation in the Social Sciences[1]) carried out by CentERdata. This project creates facilities for collecting data and focuses on experimental designs in the field of research methods in the social sciences. For this purpose a Dutch internet household panel was formed for which the participants were selected using random sampling. Households without internet access were given a set-top box with which questionnaires could be completed using a television screen as a monitor. Different from typical internet samples that do not include people without internet access, or many experimental studies that often use selective samples from homogeneous populations (*e.g.*, student populations) the MESS project guarantees a heterogeneous population of respondents. A total of 6843 panel members of 16 years or older participated in our experiment.

Questionnaire

Our questionnaire included three scales and 12 questions in total (four questions per scale, see Appendix A). Each set of questions was balanced including two positively worded and two negatively worded items. The first scale concerned attitudes toward working mothers (α values ranging between .711 and .767) and was based on the International Social Survey Program (ISSP, 2002). The second scale included issues regarding enjoyment of nature (α values ranging between .793 and .806) that were selected from the ENV scale (Bogner & Wiseman, 1999) and the Ecocentric and Anthropocentric Environmental Attitudes Scale (Thompson & Barton, 1994). The third and last scale inquired into ethnocentric attitudes (α values ranging between .795 and .816) and was adapted from the Belgian 1995 General Elections Survey (ISPO, 1997). These scales were part of a larger questionnaire designed by CentERdata and were positioned at the end of this questionnaire. The questionnaire was sent electronically to all panel members and they were asked to fill it out and send it back as soon as possible. The questionnaire was accessible during one month (January 2008) and two reminders were sent during this period. A response rate of 79.9 percent was reached within the panel (AAPOR RR6). At the offset of the MESS project 48 percent of the selected households chose to participate in the panel study.

[1]see http://www.centerdata.nl/en/TopMenu/Projecten/MESS/.

Design

At the start of our project we were aware of the fact that there are plenty of response format issues that could intervene with response styles when answering attitude questions; all of them which could not be handled in a single design. In this study a split ballot design was used in which we randomly assigned respondents to one of six conditions that only differed in the number of response options that were offered. We distinguished between shorter scale formats with 5, 6, and 7 categories, and longer formats with 9, 10, and 11 categories. Sample sizes for the 5-point and 11-point treatments were larger to allow for future experimenting in a second wave. All aspects other than scale length of the survey design were held constant for all groups and were in line with the format respondents were acquainted with in the MESS-project in which our experiment was implemented. We felt that changes in the format respondents are used to in answering the panel survey might invoke confusion and mental processes that would be difficult to distinguish from true format differences in which we were interested. Response categories were numbered and only the endpoints of the scales were labeled: 'completely disagree' for the lowest value on the left and 'completely agree' for the highest value on the right. In the MESS-panel, respondents are used to answer all questions; the program does not allow skipping questions. If respondents tried to proceed when no response was given, a message popped up requesting them to answer the question before proceeding. When respondents had decided on their response, they had to click a 'proceed' button to continue to the next question. It was not possible for respondents to return to a formerly answered question and alter their response. Also, a 'don't know' option was not included. As indicated, we are fully aware that any of these format issues might be related to response styles in their own right, but in this study we are solely interested in the effects of variations in the number of answering categories. The consequence of our choices is that we can comment on main effects of length of response scales. However, whether length of response scales interact with item non-response or 'don't know' options, cannot be documented in this study.

Method

The method employed in this study has been described in detail in Moors (2003). The model builds upon the CFA-model developed by Billiet and McClendon (2000) to control for acquiescence. By using a latent class CFA-model it was possible to diagnose ERS and—by extension—any type of response style revealing preference for particular categories of a response scale (like MRS) (Moors, 2003).

The approach that is suggested in these references is to model a confirmatory factor analysis (CFA) in which two factors are added to indicate the content of two independent sets of items (*i.e.*, the content factors), and one

additional factor is included to indicate acquiescence or ERS (*i.e.*, the style factor). Since response style guides the way respondents answer attitudinal questions, it can be thought of as a common factor that transcends independent items or theoretical concepts. Therefore, independent of item content, it should indeed be possible to identify such a response style factor within a multidimensional context (Moors, 2003).

The latent class factor approach is chosen in this research for one particular reason, *i.e.*, it allows estimating the effect of discrete interval latent class factors on nominal response variables. This in turn allows for a U-shaped relationship between response probabilities on items and the response style factor, which is inextricably bound up with ERS. In contrast, the model designed by Billiet and McClendon (2000) assumes linear relationships between these variables, making it well suited for measuring acquiescence response style, but unfit for measuring ERS. Preliminary analyses had indicated that ERS was a major concern in our data whereas little systematic evidence was found for acquiescence. Hence our choice for the latent class variant of the approach. As is demonstrated in Moors (2003) and confirmed in this research, this approach is flexible in detecting response styles related to specific response categories of the observed indicators. An extreme response style occurs when the latent class style factor reveals higher likelihoods of extreme responses relative to the other response options. A midpoint response style is revealed when the midpoint or middle response categories are relatively more chosen than the adjacent categories.

The models presented in this research included one 'style factor' influencing responses on all 12 items, and three 'content factors'—one factor for each set of items.

In equation (1) we present a simplified version of the latent class factor models that are used in this research. Assume a model with two sets of two items (A and B), two 'content' latent class factors (X_1 and X_2) and one 'style' latent class factor (X_3). Then the conditional response probabilities of this latent class factor model can be written as:

$$\prod_{k=1}^{4} \pi(A_1 A_2 B_1 B_2 | X_1 X_2 X_3) \tag{1}$$

The response probabilities of this model are restricted by means of logit models with linear terms:

$$\begin{aligned} \eta_{A_1 A_2 B_1 B_2 | X_1 X_2 X_3} = \beta^0_{A_1} + \beta^0_{A_2} + \beta^0_{B_1} + \beta^0_{B_2} + \beta^1_{A_1}.\upsilon_{X_1} + \beta^1_{A_2}.\upsilon_{X_1} + \beta^1_{B_1}.\upsilon_{X_2} \\ + \beta^1_{B_2}.\upsilon_{X_2} + \beta^2_{A_1}.\upsilon_{X_3} + \beta^2_{A_2}.\upsilon_{X_3} + \beta^2_{B_1}.\upsilon_{X_3} + \beta^2_{B_2}.\upsilon_{X_3} \end{aligned} \tag{2}$$

Since a latent class factor approach assumes that the factors are discrete interval (or ordinal) variables, the two-variable terms (*e.g.*, $\beta^1_{A_1}.\upsilon_{X_1}$) are

restricted by using fixed category scores for the different categories of the latent class factor. Equidistant scores υ_X range from 0 to 1, with the first category of a factor getting the score 0 and the last category getting the score 1. Hence, a latent class factor with, for instance, three categories gets the scores 0, 0.5, and 1. As such the categories of the factors are ordered by the use of fixed equal-interval category scores. The β's indicate the strength of the relationship between factors and response variables. Equation (2) identifies a confirmatory latent class factor model with factor X_1 influencing the response probabilities of items A_1 and A_2; factor X_2 influencing items B_1 and B_2; and factor X_3 influencing all four items. Analyses were run with LatentGold 4.5. More technical and specific details concerning latent class factor analysis are presented in Magidson and Vermunt (2001) and Moors (2003).

Conceptually, this latent class factor model is highly similar to models estimated in confirmatory factor analysis using Lisrel-type of modeling. However, there are two differences to which we would like to draw attention. First, latent class factor models involve estimating effects of the discrete level continuous latent factor on each category of the response variables. Hence, when five response categories are used, five effects (β's) are estimated. Second, the β's should not be confused with standardized effects since the method involves loglinear modeling. To facilitate the intuitive reading of our models, Figure 1 includes a graphical presentation of the model in which five response options were administered to the respondents. Similar models were estimated for the other conditions.

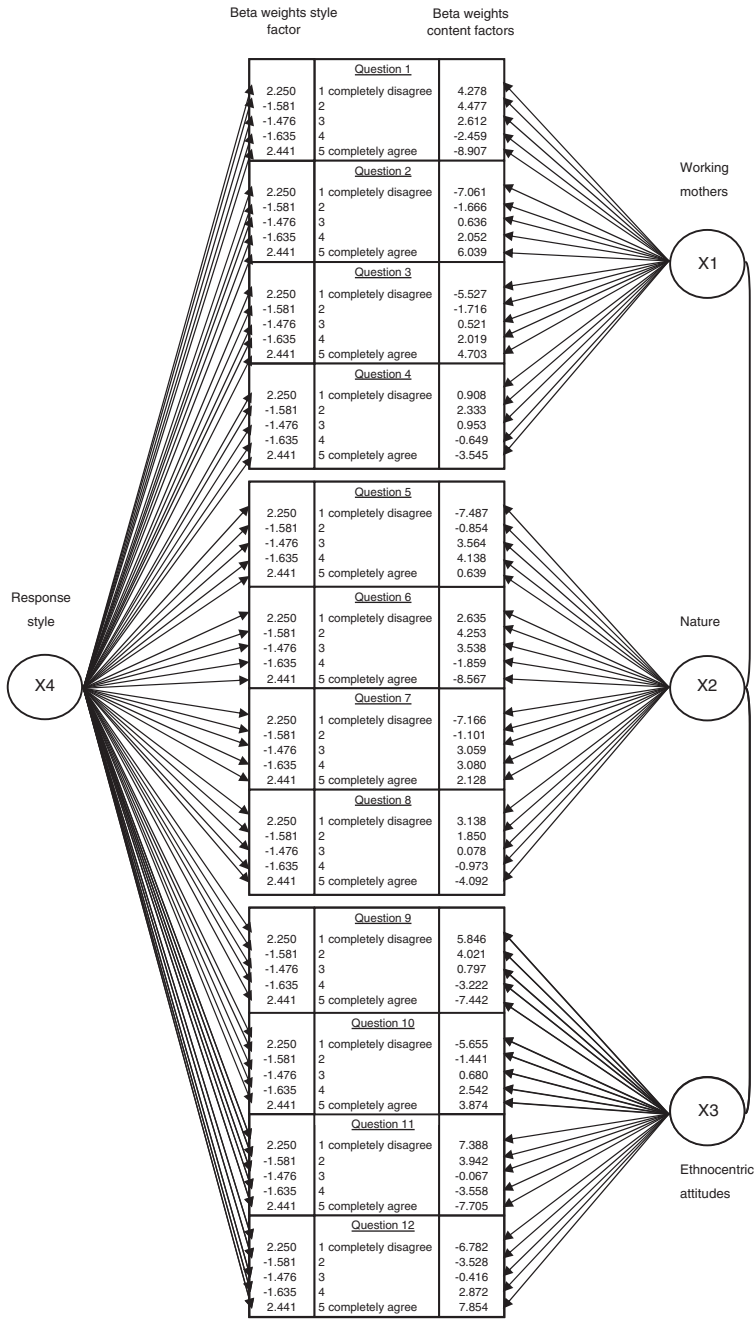
Analyses

Figure 1 represents the concept of our final models. As mentioned before, given that indicators are treated as nominal, there are just as many effects (represented by arrows) of the latent factors as there are response categories for an indicator. Other features of the model can also be read from Figure 1:

(a) The model includes three content latent class factors (X1, X2, and X3) and one style factor (X4);
(b) Content factors only influence the responses on the corresponding items, whereas the style factor is assumed to influence the responses on all items;
(c) The content factors are allowed to correlate with each other but not with the style factor; and
(d) The model imposes equality constraints in such a way that the effect of the latent class style factor (X4) is equal for all items.

The steps that led to this final model are explained in the following paragraph.

FIGURE 1 **Model of factor analysis used in our study**



To get to our final results three steps in the analysis were taken. The first step concerned deciding on the number of levels or categories the discrete interval latent class factors consist of, which determines the fixed equidistant category scores of the factors. To solve this issue, we ran our analyses several times (while varying the number of levels) for all six treatments, so that models with different numbers of levels could be compared. Decisions were made using Bayesian Information Criterion (BIC) values. Step two involved simplifying the model by imposing equality restrictions on the effects of latent class factors on response variables. Theoretically, restricting the style factor to have equal effects on all response variables made the most sense, as has also been argued by Billiet and McClendon (2000). After all, a response style is unrelated to the content of the concept that is being measured. Hence, when a response style is truly a way of answering attitude questions that respondents adopt, its effect can be assumed to be equal for all items. By comparing the unrestricted and the restricted model we check this assumption. Furthermore, we extended this effort to the content factors as well. The third step was to employ the model that proved to be optimal according to the analyses in the first two steps (like the one presented in Figure 1). Findings regarding the three steps in our analyses are reported in the next section.

RESULTS

The first step of the analysis concerned the number of levels (or discrete categories) the factors should consist of. We ran several analyses with 2, 3, 4, 5, and 6 levels on all treatments (Table 1). According to the BIC values that we found, the fit of all models improved remarkably when using 3 levels instead of 2 levels, but using 4 levels instead of 3 did not bring about sizeable improvement. Therefore we chose to include 3 levels, although it is worth noting that using either 2 or 4 levels did not alter the conclusions. Note that when we increased the number of levels, this increase was applied to the three content factors as well as the style factor.

The second step involved the simplification of models by imposing equality restrictions on the beta weights corresponding to the effects of latent class factors on the indicators. We compared the model with no restrictions with two alternative models, *i.e.*, the first with equality restrictions for the response style factor only; the second adding equality restrictions for the content factors as well. The best fitting model as indicated by BIC was a model in which equality restrictions for the response style factor were implemented, but not for the content factors (Table 2). Therefore, in this study, the effects of the response style factor on the indicators were restricted to be equal across all indicators. An example of this model is presented in Figure 1. An additional benefit of these equality restrictions is that it reduces the number of comparisons between the split samples we need to make.

TABLE 1 **BIC values of factor models with varying numbers of levels**

Model	*LL*	*BIC(LL)*	*Npar*	*df*	*p-value*
Two levels	−140790	286411	547	6293	0
Three levels	−140008	284855	548	6292	0
Four levels	−139855	284558	549	6291	0
Five levels	−139773	284403	550	6290	0
Six levels	−139738	284342	551	6289	0

TABLE 2 **BIC values of models with varying equality restrictions**

Equality restrictions	*LL*	*BIC(LL)*	*Npar*	*df*	*p-value*
No restrictions	−34413	70027.3	155	2153	0
Restrictions on style factor	−34546	69952.5	111	2197	0
Restrictions on all factors	−34931	70442.8	75	2233	0

The last step was to run the analysis that fitted our data best and compare how response styles varied according to the length of the response scales. In accordance with the results obtained by carrying out the first two steps, we ran our model choosing three levels for all factors and restricting only the style factor to have equal beta weights. To interpret how latent class factors relate to the nominal indicators we need to have a closer look at the beta weights between factors and indicators.

In Table 3 we present information regarding the response style factor. The first thing that attracts attention is that in every treatment, the beta weights corresponding to the categories at the endpoints of the scales are significantly higher than the beta weights corresponding to the categories lying in between.[2] This pattern clearly indicates the exertion of ERS, with respondents employing the extreme categories more often than other categories. Since ERS can be observed in every single treatment, there seems to be no difference in the exertion of ERS behavior when it comes to the number of response options offered—at least not as far as the formats implemented in the experiment.

Besides the length of response scales, the presence of the middle response option is another variation in response scale format that is under investigation in this study. The presence of a midpoint had no influence on the likelihood of ERS. However, we do observe some differences in beta weights that might

[2]This pattern was also observed when no equality restrictions on the style factor were imposed indicating that the findings presented in the article is not an artifact of model restrictions.

TABLE 3 **Beta weights and standard errors of the style factor for all treatments**

Number of answering categories											
5		*6*		*7*		*9*		*10*		*11*	
Beta weights	*Standard error*	*Beta weights*	*Standard error*	*Beta weights*	*Standard error*	*Beta weights*	*Standard error*	*Beta weights*	*Standard error*	*Beta weights*	*Standard error*
2.250	0.142	2.609	0.234	3.000	0.329	3.868	0.324	4.429	0.375	3.749	0.225
−1.581	0.135	−1.411	0.181	−1.390	0.348	−0.750	0.320	−0.734	0.312	−0.368	0.261
−1.476	0.141	−1.319	0.258	−1.693	0.201	−1.055	0.234	1.108	0.214	−0.863	0.189
−1.635	0.109	−1.602	0.215	−0.987	0.339	−1.578	0.238	−1.555	0.373	−1.650	0.181
2.441	0.191	−0.951	0.173	−1.277	0.220	−1.391	0.381	−0.742	0.413	−1.315	0.271
		2.673	0.277	−0.963	0.213	−2.134	0.306	−2.225	0.354	−0.938	0.301
				3.310	0.258	−1.198	0.222	−1.772	0.263	−1.645	0.186
						−0.654	0.234	−0.597	0.224	−1.157	0.154
						3.893	0.284	−0.161	0.380	−0.541	0.165
								4.462	0.348	0.318	0.285
										4.409	0.254

indicate MRS. Figures 2 and 3 allow for a clearer picture of this tendency toward MRS. In these figures the scales are rescaled in such a way that they are equal in length, so that the endpoints and midpoints of the scales can be mutually compared. In the shorter scale formats (Figure 2) there is no clear evidence that respondents disproportionately prefer midpoints over adjacent categories. However, when using a 9-point or 10-point scale (Figure 3), beta weights corresponding to the fifth category peak and, following from the confidence intervals, significantly deviate from the betas from the adjacent categories. This deviation is leveled off in the 11-point scale condition. This finding suggests two things. The first one is that shorter response scales do not evoke MRS *or* that MRS is possibly more difficult to diagnose in these types of scales. The second one is that MRS is not merely a bias caused by offering respondents a middle answer, which is indicated by the fact that the fifth response category also peaks in the 10-point scale, which has no middle option. The latter is in contrast with earlier assumptions (Kalton, Roberts, & Holt, 1980; Weijters, 2006). Our tentative—or even hypothetical—interpretation of this finding is that when encountering a 10-point scale, through lack of an exact midpoint, respondents 'create' an alternative midpoint that serves as a regular midpoint. The alternative midpoint tends to be the response category that is nearest to the middle of the scale and positioned slightly to the left. Presumably, respondents mentally divide the scale length by two to obtain the alternative middle response option of the scale (*i.e.*, 'ten divided by two equals five'). This especially makes sense since the response categories of the scales were all numbered, thus making it easy to select the equated response option. This 'division by two' principal to create an alternative midpoint is also somewhat visible in the 6-point response scale, however without being significant. We lack data to further support this interpretation, but we do feel it is well worth taking this point for future research. Following from this finding, much like in the case of ERS, the presence of a midpoint does not have influence on whether or not MRS is exerted. What is clear, is that MRS is not merely the counterpart of ERS. If that would have been the case, we would have observed MRS more clearly in each condition. Instead we observed that MRS varied as scale length varied, with little evidence of MRS in the shorter versions, a clear pattern when a 9-point or 10-point scale was used, and a less pronounced pattern when an 11-point scale was administered. At this point, we wholeheartedly admit that we need additional research to really understand why MRS popped up within the 9-point and 10-point scale conditions and not so much or not at all in other conditions. Saris' (1988) suggestion that the need for anchoring points depends on scale length is definitely a perspective worthy of attention since it can also account for our finding that respondents seem to create their own alternative midpoint in the absence of a middle response option.

FIGURE 2 **Beta weights of the style factor on rescaled categories of the (a) 5-point, (b) 6-point, and (c) 7-point treatment**

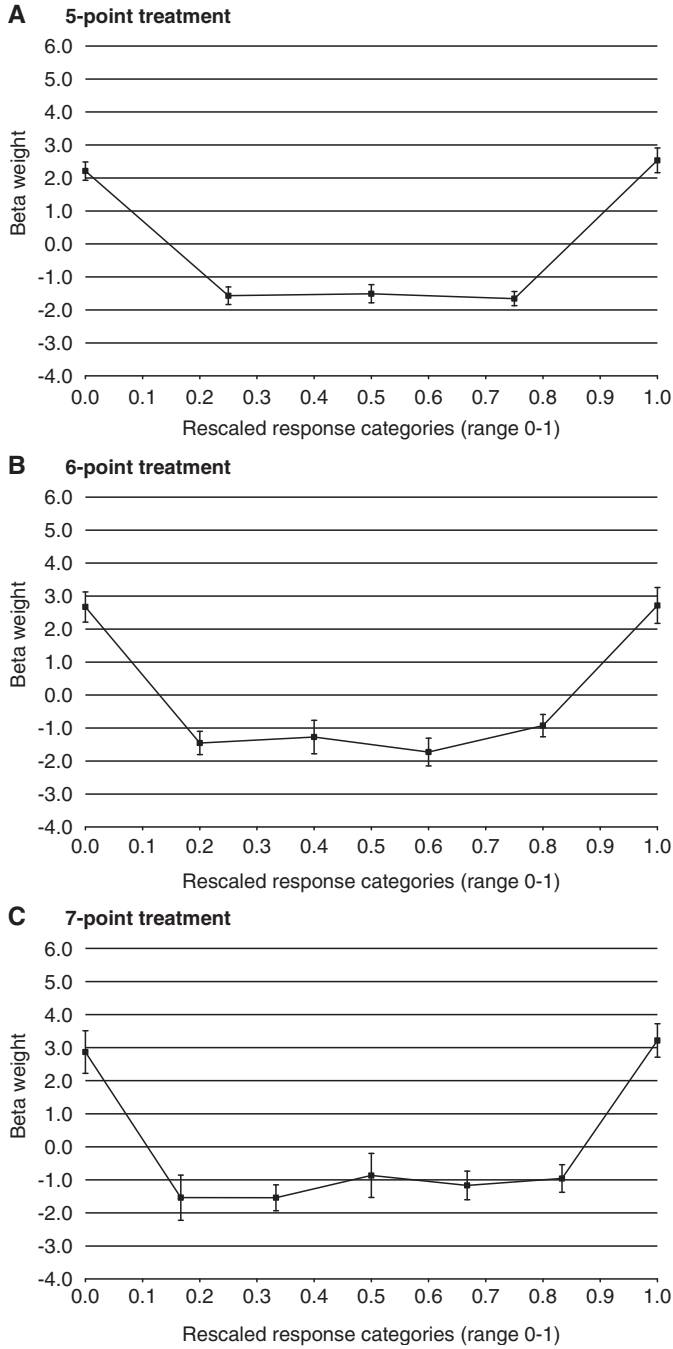
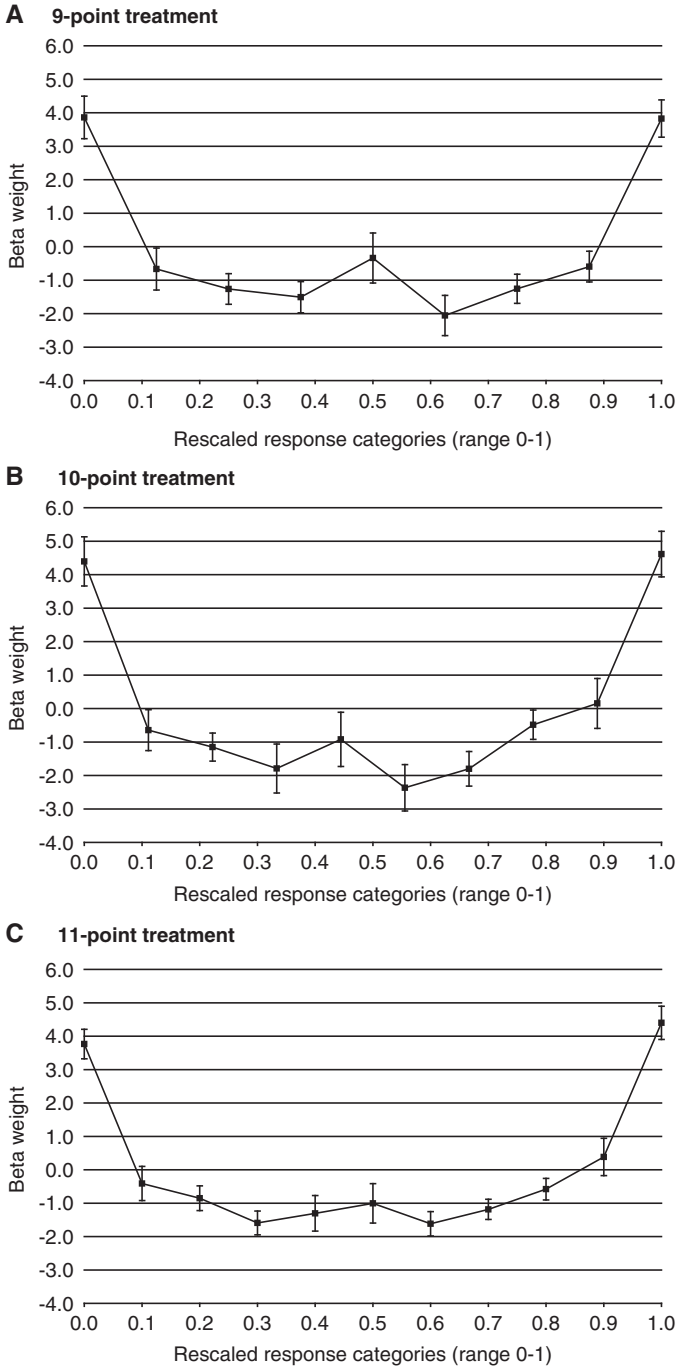


FIGURE 3 **Beta weights of the style factor on rescaled categories of the (a) 9-point, (b) 10-point, and (c) 11-point treatment**



DISCUSSION

Using a latent class confirmatory factor analysis, we have demonstrated that ERS and MRS are present in our dataset. The presence of a middle response option did not have any influence on the use of either of these response styles. The length of response scales, on the other hand, did influence (the detection of) response style behavior. However, this was only the case for the exertion of MRS, and not for ERS. MRS emerged in treatments in which a relatively large number of 9 or 10 answering categories were offered, whereas no MRS emerged if a relatively small number of answering categories was offered. As mentioned before, all treatments had similar effects on the exertion of ERS, thereby not influencing this type of response style behavior. The practical relevance of these findings are that whenever a researcher wants to control for ERS when measuring attitudes, he or she will most likely be able to detect it when response scales include the number of response categories investigated in this research. However, when MRS is a source of concern we advice the use of the longer 9-point or 10-point scale lengths.

The fact that MRS was influenced by variations in question format implies that in this study MRS was evoked by the method. ERS on the other hand was present in every single condition, suggesting that it was unrelated to the method as such. This raises the question whether ERS might be brought about by certain factors within respondents. For example, Austin, Deary, and Egan (2006) found that people high in conscientiousness and extraversion are more inclined to use ERS. Meisenberg and Williams (2008) showed that maleness is the best predictor of ERS. In addition to maleness, other predictors were associated with ERS as well, namely older age, low education and low income. Also, Baumgartner and Steenkamp (2001) found that both younger and older people tend to respond extremely. In future research we would like to link our results to socio-demographic characteristics as well as personality measures.

Some other interesting topics for future research arise from our research. For example, if the proposed strategy of creating an alternative midpoint is indeed the strategy respondents were using, then this strategy might have been different if the scale was not numbered. Numbering the scale makes it easier to create the midpoint using the method we described earlier, since respondents can make use of the numbered answering categories by reading of the alternative midpoint. When scales are not numbered and no exact midpoint is offered, one might assume that respondents will more randomly select any of the middle response options located around the midpoint and hence revealing a mild response style.

Another interesting topic for future research would be to vary different aspects of question format. In this study we varied the length of the response

scale, but numerous other factors influencing answering strategies might come to mind. For example, presenting or omitting labels corresponding to response categories of answering scales might evoke different levels of response style behavior. Also, the effect of presenting respondents with a 'don't know' option might be an interesting research topic.

While containing much strength like random assignment of respondents to treatments and the use of a latent class factor analysis to detect response style, our research also contained some limitations. First, we originally designed the study to investigate the differential effects that short scale formats and long scale formats might have on response styles. A scale with eight response categories was not included, neither were scales with more than 11 categories. The results showed that it proved to be a very sensible choice to focus on short and long versions of the same attitude scales. However, since the 8-point scale would lie in between these two sets of scales, it would have been interesting to see whether this scale has the same effect on response style as the short scales or the longer 9-point and 10-point scales in our study had. Moreover, it would have added to the evidence regarding alternative midpoints if we could have demonstrated that respondents created an alternative midpoint for the 8-point scale as well. Adding information from a 12-point or longer scale format could provide evidence whether MRS continues to be less clearly observed when increasing scale points beyond ten.

Nevertheless, the results of this study have indicated that carefully reflecting on how many response categories should be included to evoke or to be able to detect response bias, is an effort well worth making.

APPENDIX A

Overview of Items

1a) A working mother can establish just as warm and secure a relationship with her children as a mother who does not work (+).
1b) A pre-school child is likely to suffer if his or her mother works (–).
1c) All in all, family life suffers when the woman has a full-time job (–).
1d) There is more in life than a family and children, what a woman also needs is a job that satisfies her (+).
2a) I am NOT the kind of person who loves spending time in wild, untamed wilderness areas (–).
2b) I really like going on trips into the countryside, for example to forests or fields (+).
2c) I find it very boring being out in the wild countryside (–).
2d) Sometimes when I am unhappy, I find comfort in nature (+).
3a) In general, immigrants can be trusted (+).

3b) Guest workers are a threat to the employment of Dutch people (–).
3c) The presence of different cultures enriches our society (+).
3d) Muslims are a threat to our culture and customs (–).

ACKNOWLEDGEMENTS

In this article data from the LISS panel of CentERdata is used. We greatly acknowledge the contribution of the reviewers of the first draft of our article. This research was supported by a grant from the Dutch Science Foundation NWO (n° 400-06-052).

REFERENCES

Alwin, D. F. (1992). Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. *Sociological Methodology, 22*, 83–118.

Alwin, D. F. (1997). Feeling thermometers versus 7-point scales. Which are better? *Sociological Methods and Research, 25*, 318–340.

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

Arce-Ferrer, A. J., & Ketterer, J. J. (2003). The effect of scale tailoring for cross-cultural application on scale reliability and construct validity. *Educational and Psychological Measurement, 63*, 484–501.

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences, 40*, 1235–1245.

Baumgartner, H., & Steenkamp, J. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143–156.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociologicial Methods Research, 36*, 542–562.

Billiet, J. B., & McClendon, M. J. (2000). Modelling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*, 608–628.

Bogner, F. X., & Wiseman, M. (1999). Towards measuring adolescent environmental perception. *European Psychologist, 4*, 139–151.

Borgers, N., Hox, J., & Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity, 38*, 17–33.

Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: Monte Carlo investigation. *Applied Psychological Measurement, 9*, 31–36.

Clarke, I. (2001). Extreme response style in cross-cultural research. *International Marketing Review, 18*, 301–324.

Dolnicar, S., & Grün, B. (2007). Analytical robustness in cross-cultural comparisons. *International Journal of Culture, Tourism and Hospitality Research, 1*, 140–160.

Hamid, P. N., Lai, J. C. L., & Cheung, S. (2001). Response bias and public private self-consciousness in Chinese. *Social Behaviour and Personality, 29*, 733–742.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20*, 296–309.

Hurley, J. R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology, 132*, 202–210.

Institute of Social and Political Opinion Research (ISPO) (1997). *1995 General election study Belgium-Flanders* [Questionnaire and code book]. Retrieved, October 1, 2009, from http://www.data-archive.ac.uk/findingData.

International Social Survey Programme (ISSP) (2002). *Family and Gender Roles III* [Questionnaire and code book]. Retrieved, October 1, 2009, from http://www.issp.org/data.shtml.

Johnson, T. R., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*, 264–277.

Kalton, G., Roberts, J., & Holt, D. (1980). The effects of offering a middle response option with opinion questions. *The Statistician, 29*, 65–78.

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

Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science, 37*, 941–964.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, L. Decker, E. DeLeeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 141–164). New York: Wiley-Interscience.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology, 31*, 223–264.

Mandal, M. K., Ida, Y., Harizuka, S., & Upadhaya, N. (1999). Cultural difference in hand preference: Evidence from India and Japan. *International Journal of Psychology, 34*, 59–66.

Marín, G., Gamba, R. J., & Marín, B. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross Cultural Psychology, 23*, 498–509.

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44*, 1539–1550.

Miller, G. A. (1956). The magical number of seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*, 81–97.

Moors, G. (2003). Diagnosing response style behaviour by means of a latent class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination re-examined. *Quality & Quantity, 37*, 277–302.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity, 42*, 779–794.

Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences, 38*, 61–69.

O'Muircheartaigh, C., Krosnick, J. A., & Helic, A. (2000). *Middle alternatives, acquiescence, and the quality of questionnaire data*. Retrieved, October 1, 2009, from http://harrisschool.uchicago.edu/About/publications.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wright (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*, 1–15.

Raaijmakers, Q. A. W., Van Hoof, A., 't Hart, H., Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on likert-type scale items: Neutral or missing values? *International Journal of Public Opinion Research, 12*, 208–216.

Saris, W. E. (1988). *Variation in response functions: A source of measurement error in attitude research*. Amsterdam: Sociometric Research Foundation.

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

Schwartz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93–105.

Si, S. X., & Cullen, J. B. (1998). Response categories and potential cultural bias: Effects of an explicit middle point in cross-cultural surveys. *International Journal of Organizational Analysis, 6*, 218–230.

Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology, 122*, 151–156.

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology, 7*, 456–461.

Thomas, R., Uldall, B., & Krosnick, J. A. (2004). *How many are too many? Number of response categories and validity*. Retrieved, October 1, 2009, from http://www.allacademic.com/meta/p116103_index.html.

Thompson, S. C. G., & Barton, M. (1994). Ecocentric and anthropocentric attitudes toward the environment. *Journal of Environmental Psychology, 14*, 149–157.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *The Public Opinion Quarterly, 60*, 275–304.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*, 346–360.

Weathers, D., Sharma, S., & Niedrich, R. W. (2005). The impact of the number of scale points, dispositional factors, and the status quo decision heuristic on scale reliability and response accuracy. *Journal of Business Research, 58*, 1516–1524.

Weijters, B. (2006). *Response styles in consumer research* Retrieved, October 1, 2009, from http://www.feb.ugent.be/nl/index.asp.

Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*, 965–972.

BIOGRAPHICAL NOTES

Natalia Kieruj is a PhD student at the Department of Methodology and Statistics, Faculty of the Social and Behavioral Sciences, Tilburg University. Her PhD project about question format and response style behavior in attitude research is granted by NWO.

Guy Moors is assistant professor at the Department of Methodology and Statistics, Faculty of the Social and Behavioral Sciences, Tilburg University. His research interests are in the field of values research, survey research methodology, and latent class analysis.