

## Tilburg University

### Response style behavior

Kieruj, N.D.; Moors, G.B.D.

*Published in:*  
Quality & Quantity

*DOI:*  
[10.1007/s11135-011-9511-4](https://doi.org/10.1007/s11135-011-9511-4)

*Publication date:*  
2013

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
Kieruj, N. D., & Moors, G. B. D. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, 47(1), 193-211. <https://doi.org/10.1007/s11135-011-9511-4>

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Response style behavior: question format dependent or personal style?

Natalia D. Kieruj · Guy Moors

Published online: 18 August 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** In survey research, acquiescence response style/set (ARS) and extreme response style/set (ERS) may distort the measurement of attitudes. How response bias is evoked is still subject of research. A key question is whether it may be evoked by external factors (e.g. test conditions or fatigue) or whether it could be the result of internal factors (e.g. personality or social characteristics). In the first part of this study we explore whether scale length—the manipulated test condition—influences the occurrence of ERS and/or ARS, by varying scale length from 5 till 11 categories. In pursuit of this we apply a latent class factor model that allows for diagnosing and correcting for ERS and ARS simultaneously. Results show that ERS occurs regardless of scale length. Furthermore, we find only weak evidence of ARS. In a second step we check whether ERS might reflect an internal personal style by (a) linking it to external measures of ERS, and by (b) correlating it with a personality profile and socio-demographic characteristics. Results show that ERS is reasonably stable over questionnaires and that it is associated with the selected personality profile and age.

**Keywords** Extreme response style · Acquiescence response style · Attitude measurement · Personality · Format effects · Latent class factor analysis

## 1 Introduction

Response bias is a well known source of data contamination in attitude research. It refers to the situation in which a respondent's answer to survey questions is influenced by factors other than the concept that the researcher intends to measure. Several studies have shown the nontrivial influence response bias can have on the measurement of attitudes, which can

---

N. D. Kieruj (✉) · G. Moors  
Tilburg University, FSW-MTO, Room P.1.103, P.O. Box 90153, 5000 LE Tilburg, The Netherlands  
e-mail: n.d.kieruj@uvt.nl

G. Moors  
e-mail: guy.moors@uvt.nl

lead to less than accurate conclusions (Diamantopoulos et al. 2006; Dolcinar and Grün 2009; Heide and Grønhaug 1992; Moors 2003).

In the literature, a multitude of reasons are provided on the issue of what evokes response bias. However, as far as systematic response bias is concerned, we discern a tendency to ally with two perspectives that distinguish between external circumstances versus internal dispositions of respondents. First, when it comes to the former perspective, we primarily focus on those external circumstances that a researcher is able to control by design of test conditions like the format of the questionnaire (Kieruj and Moors 2010; Koson et al. 1970; Shulman 1973). Test conditions could either concern the layout or format of the questions and rating scales that are used, as well as the content or wording of questions. In the present study we designed a split-ballot research in which one particular test condition was manipulated, namely the length of the response scale. This aspect of test conditions was chosen since the ideal length of response scales is a very commonly raised issue by survey practitioners. When test conditions cause response bias, the bias is often defined as a *response set* (Naemi 2006; Rorer 1965). A second perspective is that response bias can be related to internal dispositions and characteristics of respondents (Couch and Keniston 1960; DiStefano and Motl 2009; Naemi et al. 2009). For instance, cultural and socio-demographic differences in response styles have been documented (Greenleaf 1992; Hui and Triandis 1989; Marín et al. 1992; Meisenberg and Williams 2008). Additionally, it has been suggested that response bias may be evoked by a kind of inner trait that leads respondents to systematically respond in a manner that has little or nothing to do with the questions asked (Couch and Keniston 1960). Response bias caused by internal dispositions of respondents would be called *response style* (Naemi 2006; Rorer 1965). Although we are on a thin line in disentangling the concept of ‘response sets’ from ‘response styles’, and probably not every single cause of response bias might be classified in one of these two categories, we do feel the need for a heuristic tool that—at least conceptually—distinguishes between respectively external circumstances (response set behavior) and internal characteristics (response style behavior).

The types of bias we focus on are extreme response style/set (ERS) and acquiescence response style/set (ARS). ERS is the tendency of respondents to choose the extreme end-points of a scale (Hurley 1998) and ARS is the tendency to agree rather than disagree with items, regardless of item content (Van Herk et al. 2004). The key question asked is whether ERS and ARS—i.e. the response biases of interest—are related to internal characteristics of the respondent or whether they are the result of external properties of test conditions which can be manipulated by the researcher. The former will be investigated by (a) checking if certain personality traits are related to ERS or ARS, by (b) checking how consistent the use of ERS and ARS by respondents is across different questionnaires, and by (c) investigating whether certain demographic values are related to the use of ERS and/or ARS. Whether or not ERS and ARS are the result of external properties will be investigated by checking if varying the length of the response scale will influence the use of ERS and ARS.

We have several reasons to focus on ERS. In our own experience, it tends to be present in a lot of datasets. Whenever this is the case and a particular response bias is recurrently observed, it might indicate that it is consistently used over questionnaires and might therefore be an expression of a personality trait. Furthermore it is found that ERS tends to differ across cultures (which can be seen as a stable respondent characteristic) (Chen et al. 1995; Dolcinar and Grün 2007; Hui and Triandis 1989; Johnson et al. 2005; Marín et al. 1992; Van Herk et al. 2004) and variables like age and level of education or income also play a role in the employment of ERS (Meisenberg and Williams 2008). The choice for ARS was made, because like ERS this is one of the two most commonly discussed response biases in attitude

research. Like ERS it also has often been linked to culture (Cheung and Rensvold 2000; Johnson et al. 2005; Marín et al. 1992; Smith 2004), and incidentally, ARS has also been discussed in relation to certain personality traits (Couch and Keniston 1960). Furthermore, two of the three sets of questions used in this research come from a study in which ARS was found in a lengthy face-to-face survey (Billiet and McClendon 2000). Using panel data including the same questions, Billiet and Davidov (2008) have argued that the persistence of ARS across waves indicates that ARS is a personality trait.

Our research adds to the existing literature by attempting to bring more clarity to the origin of ERS and ARS. Whether the origin of these response biases lies within the individual or whether it is the result of certain test conditions is an important question for attitude research. The answer has implications for the best way of dealing with ERS and ARS. If response bias were solely a result of test conditions and thus were a response set, it would be interesting to figure out what is evoking response bias and how to prevent it from occurring. Although ambitious since probably more than one test related factor could play a role, the regular survey practitioner would be enthusiastic if he or she could minimize the problem by adopting an appropriate survey design. In this research, which focuses on the length of the response scales used in attitude research, this would imply that we would be able to identify the optimum number of response categories and as a consequence minimize the occurrence of response bias. However, to the extent that response bias is influenced by properties of the individual, it is probably not possible to prevent them from occurring. If so, a “preventive check” by adopting a particular design is less useful and hence the need to correct for response bias in measurement models increases. Therefore, as an important secondary goal of this study, we present an extension to the method that builds upon the latent class confirmatory factor model that has been recently introduced for diagnosing and controlling for ERS (Moors 2003; Morren 2011) in which ARS can be diagnosed simultaneously.

The paper is organized as follows. Firstly, we give an overview of the existing literature about response bias being the result of test conditions and about response bias being a person-related bias. Secondly, we introduce the latent class confirmatory factor model that allows for diagnosing ERS and ARS simultaneously. Given that this approach has only been recently developed and that we extend the approach to account for two types of response styles instead of one, we devote ample attention to it. Thirdly, we explore whether ERS and ARS are linked to the test condition of interest, i.e. the length of the response scale, and examine to what extent ERS and ARS are related to certain personality traits. Also, we investigate to what extent they are consistent over questionnaires and whether or not they are related to demographic values. Finally, results and conclusions are reported.

## **2 Is response bias the result of respondent characteristics or test conditions and circumstances? A literature review**

We would like to emphasize the significance of the question raised in this title. Any survey researcher's goal is to establish valid and reliable measurements of the concepts central to his or her research question. Therefore it is important to make sure responses are as unbiased as possible. As stated before, it is our opinion that response bias can be prevented to the extent that factors external to the individual constitute the primary source of response bias. When characteristics of the individual are involved, however, the need for a statistical ‘cure’ to control for response bias comes to the fore.

## 2.1 The length of response scales as a test condition

In this research we vary the length of the response scale to check if this property of response scales has an influence on response bias. A couple of reasons why the length of response scales could have such an influence come to mind. One of these reasons is that longer response scales might lead to increased task difficulty compared to shorter scales. If the use of a particular scale is too strenuous, respondents might become frustrated and as a result lose motivation to give accurate responses. In such a case, respondents might use a heuristic to fill out the questionnaire without having to actually process the questions. This phenomenon has been described by [Krosnick \(1991\)](#) as satisficing. Krosnick states that respondents, who are not willing to expand the necessary effort and time to form optimal answers to attitude questions, might adopt heuristic shortcuts to formulate answers that satisfy them enough.

Scale length can also be directly linked to scale sensitivity. As a scale becomes longer it naturally becomes more sensitive, and respondents can indicate their opinion with more precision than when shorter scales are used. At a certain point however, adding answering categories could lead to confusion on the respondents' side, since it might not be clear what the difference is between two neighboring categories. The confusion might lead to 'satisficing' in the form of ERS or ARS.

In this study, one of the goals is to find the response scale length that has the exact right sensitivity (and thus length) to it.

## 2.2 Person related characteristics

If test conditions do not affect response style behavior, the idea that stable components within the individual like personality characteristics influence response behavior gains momentum. As [Hamilton \(1968\)](#) stated, the evidence that ERS scores are reliable suggests that this response tendency may have personality concomitants of its own. In fact, [Greenleaf's model \(1992\)](#) for measuring ERS, for instance, rests entirely upon the idea that ERS is an individual trait. Basically, [Greenleaf](#) argues that it is possible to compute an ERS index by counting the extreme responses to a large set of conceptually unrelated items and using this to correct for extreme response bias in any given model including attitudes. As such, he defines an external measure for ERS. This procedure has two implications. First, since [Greenleaf](#) selects items intended to measure different concepts, it is implied by definition that ERS occurs across all kinds of attitudes. Second, [Greenleaf](#) does not explicitly consider ERS to depend on the length of the response scale that is used. In this sense, his method to correct for response bias entirely rests upon the assumption of response bias as a personality trait.

Several researchers have tried to link ERS to personality characteristics. For example [Austin et al. \(2006\)](#) found that extreme responders are more likely to be extraverted and high in conscientiousness than subjects who are not extreme responders. Also, [Lewis and Taylor \(1955\)](#) found that respondents with high anxiety scores use ERS more often than respondents low in anxiety. An interaction effect was found by [Naemi et al. \(2009\)](#), implying that those who finish a questionnaire quickly and those who are either tolerant of ambiguity or are simplistic thinkers are most likely to exhibit ERS. In another study by [Naemi \(2006\)](#) it was found that a positive relationship exists between intolerance of ambiguity and ERS, and that respondents who finish a survey quickly and score high on decisiveness are more likely to exert ERS. Only a few studies have been dedicated to identifying personality correlates of ARS. In one study, [Pedersen \(1967\)](#) found that rathymia (the state of being cheerful, merry and optimistic) is positively correlated to ARS and that cooperativeness is negatively

correlated to ARS. Another study by Knowles and Nathan (1997) showed that ARS is related to cognitive simplicity, rigid mental organization and intolerance of alternatives.

Other person-related factors that have been linked to response style refer to socio-demographic characteristics such as age, income and education. For example, Meisenberg and Williams (2008) showed that ERS and ARS are both positively related to age and negatively related to education and income. In accordance with these findings, Greenleaf (1992) found that age, income and education influence respondents' exertion of ERS.

In the current study we selected personality scales to form a combination of personality traits (personality profile) that was expected to be relevant to the use of ERS and ARS, namely extraversion, agreeableness, indifference, valuing strong opinions, relational skills, black and white thinking and being intelligent/intellectual. Rather than investigating their separate effects on response bias, we have chosen to combine personality characteristics into a kind of personality profile. We think that response bias is most likely the result of a very complex combination of factors that might all contribute to some extent to the use of ERS and ARS. Therefore, we decided to take a closer look at the aforementioned personality profile, which we expect to be associated with the response biases to a certain extent.

### 2.3 Developing the research question

Given the previous arguments (concerning test conditions vs. respondent characteristics) it is difficult to formulate straightforward hypotheses regarding the occurrence of ERS and ARS. At least to some extent we expect ERS and ARS to reside within the individual and—as a consequence—we expect it to be observed regardless of scale length. However, we also presented arguments indicating that the extent to which response bias is revealed might depend on scale length (or scale sensitivity). We are fully aware of the possibility of other external factors playing a role in the employment of ERS and/or ARS. For instance, to some extent, certain properties of response scales and wording of questions undoubtedly all influence the use of these response biases as well. However, we believe that the origin of a considerable part of ERS and ARS lies within the respondent himself and to a lesser extent within the properties of the scales being used.

We will test this expectation in different ways. First, we will compare respondents' scores on ERS and ARS across questionnaires that differ in the length of the response scale used. Second, we will test whether respondents sharing a combination of personality characteristics (extraversion, agreeableness, indifference, valuing strong opinions, relational skills, black and white thinking and being intelligent/intellectual) are more likely to use ERS or ARS than respondents who do not share these characteristics. Also, we investigated whether ERS and/or ARS within our dataset can be linked to external measures of response style behavior. Lastly, we investigate if there is a link between ERS and/or ARS and socio-demographic characteristics, like age, gender, income and education level.

## 3 Data and method

### 3.1 Participants

Our questionnaire was distributed by CentERdata to their LISS panel, which is a household panel consisting of 8,044 participants in total. Different from voluntary internet panels, this panel used a random sampling design in which households were contacted personally to

participate, similar to sampling procedures used in face-to-face surveys. Panel members that did not have a personal computer or internet during the selection process were presented with a television box and/or an internet connection. Questionnaires were accessed electronically via internet providing flexibility on when respondents entered their responses. Our attitudinal scales were part of a questionnaire that was filled out by 6843 panel members in January 2008 which resulted in a response rate of 79.9% (AAPOR RR6). Of the respondents 46.1% were men and 53.9% were women. Age ranged from 16 to 94 (mean age was 45.46).

### 3.2 Questionnaire

The methodology used in this paper, which is explained in some detail afterwards, involves the use of multiple sets of questions measuring separate concepts. Our questionnaire consisted of three sets of four items forming three different scales (Appendix A). The first scale inquired after attitudes towards working mothers ( $\alpha$ 's ranging from 0.711 to 0.767) and the items were selected from the International Social Survey Program (ISSP 2002). The second scale was about attitudes towards nature ( $\alpha$ 's ranging from 0.793 to 0.806) and was based on the ENV scale (Bogner and Wiseman 1999) and the Ecocentric and Anthropocentric Environmental Attitudes Scale (Thompson and Barton 1994). The third scale was about ethnocentric attitudes ( $\alpha$ 's ranging from 0.795 to 0.816) and items were selected from the Belgian 1995 General Elections Survey (ISPO 1997). All scales were fully balanced, meaning that they consisted of two negatively worded items and two positively worded items. This is an important criterion for diagnosing ARS (Billiet and McClendon 2000), since we can only be sure if respondents agree with all questions regardless of the content if they agree with questions inquiring after the same attitude that differ only in their valence (positive vs. negative). Our questions were added at the end of a survey and filling out the entire questionnaire took participants about 20 min. Respondents were asked to submit this questionnaire within a month and during this period, three reminders were sent by e-mail.

Since our project was included in a panel study we were able to select a number of items from previous waves that were relevant to our research. A selection of 23 items was made referring to seven personality traits which we expected to be related to ERS and ARS, i.e. extraversion, agreeableness, indifference, valuing strong opinions, relational skills, black and white thinking and being intelligent/intellectual. A second set of 18 items was selected from a pool of attitudinal questions following the methodology suggested by Greenleaf (1992) to equate a 'contentless' measure of ERS, which involves the summing of extreme responses on a set of items with low inter-item correlations. That way an independent 'external' measurement of ERS is developed. Details on this procedure are provided further on in this research.

### 3.3 Design

We implemented a split ballot design in which respondents were randomly assigned to six experimental groups that differed in the number of response scale categories presented to them. The sample size in the 5- and 11-point treatment was higher (2,309 in the 5-point group and 1,481 in the 11-point group) than in the other groups (ranging from 724 to 799 respondents) to account for future analyses. A distinction was made between short scales and long scales, with 5-, 6- and 7-point scales representing the shorter scales, and 9-, 10- and 11-point scales representing the longer scales. Response categories were all numbered, but



only the endpoints of the scales were labelled ('totally agree' on the left and 'totally disagree' on the right).

It is part of the standard procedure of the LISS-project not to offer a 'don't know' answering option, and also not to let respondents skip questions (a message would pop up, informing respondents that they could only proceed to the remainder of the questionnaire if the question is answered). We decided not to deviate from this procedure since the respondents were acquainted with it and we wanted to avoid raising suspicion regarding the experiment. We acknowledge that some of these aspects that we fixed to be equal across groups deserve attention in research on response styles. However, incorporating all these different aspects within a single design is not feasible. The main reason to focus on the number of response categories as a test condition is its practical relevance to scholars and applied researchers who like to know what the consequences are of choosing a particular number of response categories.

### 3.4 Method

Several approaches have been suggested for measuring ERS and ARS. These approaches can typically be classified as two types of measurement, namely methods in which a sum score index is constructed for extreme or acquiescent responses (Gibbons et al. 1999; Greenleaf 1992; Harzing 2006; Johnson et al. 2005; Shulman 1973), and methods that are based on statistically modelling these response biases (Billiet and McClendon 2000; Bolt and Johnson 2009; De Jong et al. 2008; Van Rosmalen et al. 2007). In this study, a model based approach is our primary method, but the sum score method will also be used as a way of cross-validating the results of the method we advocate. Examples of model based approaches include IRT (Bolt and Johnson 2009; De Jong et al. 2008), latent class (Van Rosmalen et al. 2007) or confirmatory factor analysis (Billiet and McClendon 2000). We chose a latent class confirmatory factor model in which ARS and ERS can be simultaneously modelled alongside the content factors within one single model. A major advantage of this method is that it is able to deal with the nonlinear relationship between the latent ERS factor and the manifest response items. Respondents high in ERS are expected to use both endpoint categories more often than the categories lying in between, thus resulting in a regression graph that follows a C-shaped form (describing the effect of the latent ERS factor on the response items). As demonstrated before (Moors 2003; Morren 2011) a latent-class factor analysis allows for estimating such an effect since it allows for defining the response items as nominal response variables. Consequently, separate beta weights for each answering category are estimated, thereby making it possible to reveal C-shaped relationships that are typical for ERS. In fact, this method allows for detecting other types of scale point preference being measured as well (Kieruj and Moors 2010).

In the case of ARS, we expected a linear relationship between the latent ARS factor and the response items, since we expect respondents who are high in ARS to prefer the second answering category over the first, to prefer the third category over the second, et cetera and to finally prefer the last answering category over the second last answering category (since respondents high in ARS will always prefer an answering category that is higher in agreement). Since we expect a linear relationship for ARS, the response items can be defined ordinally in this case.

Our model is based on that of Moors (2003) which was further extended by Morren (2011) showing that the effects of latent factors can be defined nominally and ordinally within the same model, depending on the type of relation that is expected between the latent factors and



the item responses. We chose to define the effects of the ERS factor to be nominal and the effects of the ARS factor (as well as the content factors) to be ordinal. Given that our study involves the use of three sets of items the multinomial logit regression equation modelling ERS as well as ARS has the following form:

$$P(Y_{ij} = c | F_{1i}, F_{2i}, F_{3i}, A_i, E_i) = \frac{\exp(\beta_{0jc} + \beta_{1jc}F_{1i} + \beta_{2jc}F_{2i} + \beta_{3jc}F_{3i} + \beta_{4jc}A_i + \beta_{5jc}E_i)}{\sum_{d=1}^C \exp(\beta_{0jd} + \beta_{1jd}F_{1i} + \beta_{2jd}F_{2i} + \beta_{3jd}F_{3i} + \beta_{4jd}A_i + \beta_{5jd}E_i)} \quad (1)$$

where  $Y_{ij}$  denotes the response of individual  $i$  on item  $j$ , and  $c$  denotes a score on a specific answering category out of  $C$  categories. The responses of individual  $i$  on the three latent content factors are represented by  $F_{1i}$ ,  $F_{2i}$  and  $F_{3i}$ . The responses of  $i$  on the ERS style factor are represented by  $E_i$  and the responses on the ARS style factor are represented by  $A_i$ .  $\beta_{0jc}$  is the intercept and  $\beta_{1j}$ ,  $\beta_{2j}$ ,  $\beta_{3j}$ ,  $\beta_{4j}$  and  $\beta_{5jc}$  are the regression coefficients, which indicate the strength of the relationship between the latent factors and the response variables. The subscript  $j$  denotes the fact that separate regression coefficients are equated for every item. Only in the case of ERS with corresponding beta weight  $\beta_{5jc}$  there is a subscript  $c$  which denotes the fact that separate beta weights are equated for every response category implying that indicators are defined nominally. The other regression coefficients (e.g.  $\beta_{3j}$ ), however, do not have this subscript  $c$  but instead are multiplied by the value of the answering category ( $c$ ) of interest, indicating that distances between answering categories are assumed to be equal and a monotone relationship between content factors and item responses exists. In other words, in all cases except that of ERS, indicators were defined ordinally. A schematic overview of the model is presented in Fig. 1, with separate effects (nominal) going from the ERS factor to the response items and with a single effect (ordinal) going from the ARS and content factors to the response items (Fig. 1 represents the 5-point treatment).

In our model, response items only loaded on their corresponding content factors, meaning that items X1–X4 loaded exclusively on  $F_{1i}$ , items X5–X8 only loaded on  $F_{2i}$ , and items X9–X12 only loaded on  $F_{3i}$ . Therefore, regression coefficients representing other combinations were restricted to be zero. However,  $A_i$  and  $E_i$  loaded on all items, since response styles are expected to affect all items (Fig. 1).

This latent class factor model defines the latent factors as ordinal discrete level variables. In a first step, there should be decided on what the number of ordered categories will be that these latent variables consist of. In this study, the discrete latent factors ( $F_{1i}$ ,  $F_{2i}$ ,  $F_{3i}$ ,  $A_i$ ,  $E_i$ ) consist of three equidistant levels, which are labeled 0, 0.5 and 1. This number of levels was chosen after comparison of models with 2, 3, 4, 5, and 6 levels per latent factor. All models were estimated with LatentGold 4.5 (<http://www.statisticalinnovations.com>). Table 1 presents information regarding 5-point scales, but results are similar for other conditions. We found that the fit of all models, defined by the Bayesian Information Criterion (BIC) based on the model's log likelihood (LL), improved remarkably when using three levels instead of 2. Model fit continues to slightly improve by increasing levels. However, given that the results from the 4-levels analyses did not substantially differ from the 3-levels analyses and that increasing the levels of the latent class factors increases computational time exponentially, we decided on using discrete latent class factors with 3 levels.

We already indicated that we reduced the complexity of the discrete latent factor model with nominal response variables by imposing ordinal restrictions in the relationship of these response variables with the content factors as well as the ARS factor. We further explored the opportunity of reducing the complexity of the model by fixing the regression coefficients

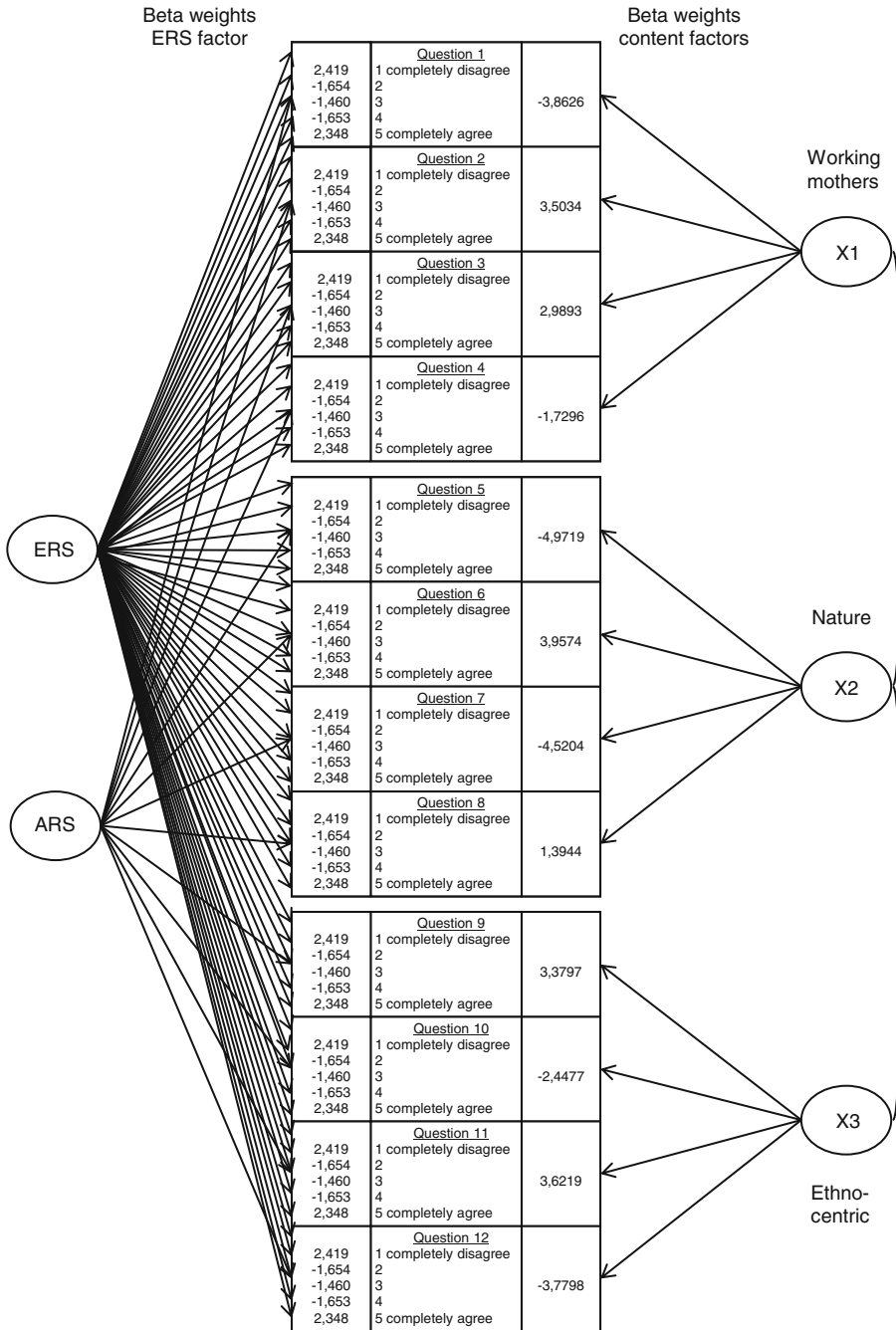


Fig. 1 Latent class confirmatory factor model with two style factors and three content factors

**Table 1** BIC values of LC factor models with varying numbers of discrete levels

Number of equidistant categories (levels)	
Model	BIC(LL)
2 levels	286,411
3 levels	284,855
4 levels	284,558
5 levels	284,403
6 levels	284,342

Results are from the 5 scale point conditions

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

Equality restrictions	BIC (LL)
No restrictions	70,027
Restrictions on style factors	69,952
Restrictions on all factors	70,443

Results are from the 5 scale point conditions

of the ERS and ARS style factor to be equal for all items. Comparing models with equality constraints on (a) all factors, (b) the style factors only, and (c) no restrictions, using BIC indicated that a model with equality restriction on style factors only (model b) fitted our data best. In Table 2 we report the typical results which were similar for all treatments using the 5-point scale treatment as an example.

### 3.5 Analyses

The following research strategy is implemented in this research. First, we test whether ERS and ARS is present in the given datasets. Second, we investigate to what extent response styles can be linked to the number of response categories. Third, we adopt the ‘contentless measure’ strategy as suggested by Greenleaf and calculate an ERS index using other items from previous waves to investigate consistency in response style used across items. Finally, we link response styles to personality traits and demographic values. The main purpose of this and the former step in the analysis is to investigate whether the use of response style merely indicates a systematic form of ‘noise’ in response patterns or rather a more ‘substantive’ characteristic of individuals.

#### 3.5.1 External measure of response style

We chose to compare results of Greenleaf’s (1992) ‘contentless’ measure performed on an external dataset to results from our latent class factor model performed on the original dataset (working mothers/nature/ethnocentric attitudes). This way it becomes possible to check whether ERS is a stable component across different format designs and item content. To find a set of items with low inter-item correlations we conducted an exploratory principal component analysis on a set of 121 attitude items included in previous waves from the LISS panel surveys. By picking items that loaded on different components we were able to identify a set of 18 items of which all of the inter-item correlations were lower than  $|\cdot 30|$ . The ‘contentless’ measure was then calculated by assigning score 1 to extreme responses and score 0 to non-extreme responses on the items, and defining a sum score that indicates the number of

times an extreme response was given. This sum score of extreme responses on a set of items external to our measurement was then linked to the internal measurement of ERS with our latent class factor models.

The basic argument of Greenleaf to select a large set of items with low inter-item correlations is that extreme responses to a set of related items might just indicate a ‘true’ extreme position on the topic. By selecting items with low inter-item correlations, the chance that certain respondents have a genuine extreme opinion on all items becomes very small. It is a small step to think of a similar index for ARS. In that case we would need a set of positively worded questions with low inter-item correlations and for each of these items a negatively worded equivalent. As indicated before, the latter is required because only agreement with both a positively and negatively worded item of the same content can be interpreted as agreement bias. Unfortunately, the LISS questionnaires do not include such items, and by consequence this analysis is restricted to ERS. Running ahead of our analyses, we like to indicate here that there was only weak evidence of ARS in our models. Hence, the need for a similar ‘contentless’ measurement of ARS was less urgent.

### 3.5.2 *Personality correlates*

One of the waves of the LISS panel prior to our experiments included personality measures, which allows us to test the hypothesis that ERS and ARS are associated to certain personality traits. As discussed before, we are interested in identifying a kind of personality syndrome or a combination of personality traits that together can predict the occurrence of ERS and/or ARS. A classic latent class model, which might be considered as the structural equation equivalent to cluster analysis, suits this purpose. We estimated such a latent class cluster model on the 23 items belonging to the seven personality scales that we expected to be relevant to ERS and ARS. The seven scales of interest were extraversion, agreeableness, indifference, valuing strong opinions, relational skills, black and white thinking and being intelligent/intellectual. Two latent-class clusters emerged from that analysis and the latent class probabilities of belonging to one of the two classes was correlated to the latent class probability scores on the ERS and ARS factor in each of the experimental settings we were investigating. Details on the meaning of these two clusters are provided in the results section.

### 3.5.3 *Demographic measures*

Finally, to check whether certain demographic measures (which can also be considered as stable components in individuals like personality traits) influence the use of ERS and ARS, we correlated the probability scores of respondents on the ERS and ARS factors to age, gender, income and education level. Respondents were asked to fill out their age, gender, net and gross income and their highest completed level of education. Net and gross income was measured in categories ranging from no “income” to “more than 7,500 euro a month” with categories increasing with 500 euro per category. Respondents were given the option to choose a “don’t know” or “don’t want to say” category. Education level was also measured in categories, ranging from “elementary school” to “university”.

## 4 Results

The same procedure is adopted for each of the six subsamples of our split-ballot experiment. Recall that these samples answered to the same set of questions and that they only differed in

**Table 3** BIC values of models varying in the composition of style factors for all treatments

Number of response categories	Short scale format			Long scale format		
	5	6	7	9	10	11
Model 1: content (no style factors)	56,221	20,839	22,617	27,122	28,565	55,294
Model 2: content + ARS	56,214	20,798	22,620	27,131	28,549	55,255
Model 3: content + ERS	55,050	20,294	22,016	26,490	27,695	53,759
Model 4: content + ARS + ERS	55,039	20,284	22,004	26,472	27,680	53,722

**Table 4** Classification statistics (standard  $R^2$ )

Number of response categories	Latent class factors				
	Gender roles	Environment	Ethnocentrism	ERS	ARS
5	0.78	0.79	0.80	0.68	0.17
6	0.80	0.83	0.83	0.73	0.19
7	0.82	0.81	0.79	0.77	0.33
9	0.79	0.83	0.84	0.74	0.33
10	0.83	0.82	0.81	0.78	0.27
11	0.81	0.81	0.82	0.78	0.26

the number of response categories that were administered, namely ‘short’ (5–7) and ‘long’ (9–11) response scales.

The first question that needs to be answered is whether ERS and ARS should be included in the measurement models. This question is answered by comparing the fit of models with and without these latent style factors based on their BIC values. We compared four models (Table 3): (a) model 1 with only three content factors; (b) two models that included one of the two style factors, ARS (model 2) or ERS (model 3); and (c) model 4 that included both ERS and ARS. Including ERS definitely improved model fit, whereas ARS only marginally improved model fit in some of the models. In each test condition, the best fitting model is the model which includes both ERS and ARS. However, the reduction in BIC compared to models that only include ERS is small. Furthermore, classification statistics revealed poor performance as far as ARS is concerned. These statistics indicate how well the model predicts the latent class factor scores of cases and are defined by a standard R-squared estimate for each latent class factor (Table 4). Whereas the standard R-squared estimates for the content factors in each model are about equal to 0.80 and for ERS on average equal to 0.75, the estimates in the case of ARS are at best equal to 0.33. Given these findings, it is best to conclude that the data reveals clear evidence of ERS, whereas ARS is only weakly observed.

#### 4.1 The effect of scale length on ERS and ARS

In Table 5 the beta weights of the ERS and ARS style factor on the items are shown. The effect of these response style factors on the response variables was restricted to be equal, thus the estimated effects presented in Table 5 apply to all 12 items. Both style factors are discrete ordinal latent factors, but they differ in how they influence the response variables. In the

case of ERS, the response variables are defined as nominal and, hence, separate beta weights are estimated for every response category. The response variables are treated as ordinal in the case of ARS and as a consequence a single beta weight defines the relationship between them and the ARS factor. The most striking finding in Table 5 is that the beta weights of the endpoints of each scale are significantly higher than the betas of the intermediate categories. This finding is consistent with the interpretation of ERS, which implies higher probabilities of choosing either extreme category compared to other scale points. Since ERS is present in all treatments, no effect of the number of response options on ERS is found.

We have indicated before that we found weak evidence of ARS. Nevertheless, in each model a significant positive beta estimate is observed which is consistent with the acquiescence response interpretation since the same beta applies for both positively and negatively worded items.

#### 4.2 External validation ERS

The latent class factor approach allowed us to diagnose the use of ERS if responding to Likert type scales of various lengths. The interpretation is inferred from the particular response pattern to the items from three different concepts. As such it should be regarded as an internal or test-specific measurement of ERS. Given that we observed ERS in each test condition, it becomes worthwhile to check whether our latent-class factor measurement of ERS can be validated by correlating it with an external measure of ERS. As explained before, we have developed a 'contentless' measure of ERS following the procedure sketched by Greenleaf and calculated a sum score of extreme responses on an external set of 18 items with low inter item correlations. We then calculated correlations of the estimated latent class ERS factor scores with the latter 'contentless' sum score of extreme responses. All correlations were significant and ranged from 0.371 to 0.493 (separate correlations were equated for each treatment). These correlations may be regarded as fairly high if we take into account that we selected items from other questionnaires administered at previous waves of the panel research and given that other items than the ones defined in our experiment were used.

#### 4.3 Personality correlates of ERS

So far our analyses have indicated that ERS is present in each test condition—hence invariant to the length of the response scale—and can be linked to an external alternative measure of ERS—hence revealing consistency across occasion and item content. Does this indicate that ERS can be interpreted as a kind of personality trait; a characteristic of an individual that repeatedly might pop up when answering survey questions? To explore the data on this topic we linked ERS to particular personality profiles. As indicated before, these profiles were defined by applying a standard latent class analysis on a set of 23 questions from seven personality scales. This analysis revealed two latent clusters; one cluster that could be interpreted as grouping individuals that are sociable, extraverted, value strong opinions and view themselves as intelligent/intellectual (for a complete overview see Appendix B) and a second cluster that could be interpreted as grouping respondents that are introverted, indifferent, not sociable and do not view themselves as intelligent/intellectual. We found that membership to the first cluster was positively associated with the ERS factor, whereas membership to the second cluster was negatively associated with the ERS factor (with Pearson correlations ranging from  $|.211|$  to  $|.349|$ ,  $p$ -values  $<0.05$ ).

**Table 5**  $\beta$  weights and standard errors of the style factors for all treatments

Response style	Treatment											
	5 cat		6 cat		7 cat		9 cat		10 cat		11 cat	
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE
ERS	2.419	0.127	2.594	0.230	3.626	0.242	3.575	0.231	4.195	0.279	3.938	0.205
	-1.654	0.088	-1.515	0.153	-1.329	0.182	-0.414	0.197	-0.167	0.183	0.034	0.166
	-1.460	0.101	-1.566	0.174	-1.777	0.186	-0.993	0.173	-0.853	0.146	-0.733	0.114
	-1.653	0.087	-1.443	0.153	-1.055	0.188	-1.331	0.205	-1.208	0.166	-1.503	0.142
	2.348	0.117	-0.825	0.133	-1.571	0.169	-0.573	0.182	-0.879	0.158	-1.422	0.147
			2.755	0.185	-1.126	0.160	-1.826	0.188	-2.102	0.165	-1.100	0.124
					3.231	0.235	-1.220	0.152	-1.439	0.131	-1.532	0.127
							-0.551	0.183	-0.651	0.123	-1.017	0.114
							3.333	0.215	-0.346	0.183	-0.608	0.113
									3.451	0.232	0.156	0.158
ARS	0.392	0.100	0.441	0.243	0.433	0.091	0.581	0.105	0.436	0.101	0.334	0.074



#### 4.4 ARS and personality

Regardless of the fact that our data only revealed weak evidence of ARS, we were proceeded to calculate the correlation of the ARS-scores with the personality cluster memberships.

However it did not come as a surprise that no sizable association was found in any of the test conditions. Naturally, we cannot be certain whether this result would be similar if stronger evidence of ARS would be found in a given dataset. Therefore we do not want to speculate as to whether or not ARS might be related to factors within the individual.

#### 4.5 ERS and ARS and socio-demographic measures

Like personality traits, socio-demographic measures can also be seen as more or less stable components, and could therefore also be related to ERS being stable over scale length and questionnaires. Age, gender, net and gross income and education level were correlated to respondents' probability scores on the ERS and ARS factor. Only age turned out to produce mentionable correlations with the response style factors (with Pearson correlations ranging from  $|0.093|$  to  $|.259|$ ,  $p$ -values  $<0.05$ ). The older respondents were, the more they were inclined to use ERS. Gender, net and gross income, and education level all produced non-significant correlations with ERS and ARS.

### 5 Discussion

We explored whether the occurrence of two typical response styles/sets, i.e. ERS and ARS, depended on the length of response scales; and if so, which scale format 'suffered' the least from the two response biases. We found no evidence that could lead to a suggestion regarding the optimum number of response categories in terms of ERS and/or ARS. Instead, strong evidence was found of ERS in each test condition, whereas ARS turned out to be of little concern. Our research merely demonstrates that ERS cannot be prevented by choosing a particular number of response categories. This does not necessarily imply that ERS cannot be prevented since other questionnaire design features were not tested. However, we do believe it will be difficult to avoid this type of response bias, since our results suggest that ERS might be more of a *response style* than a *response set*. Two analyses provided supportive evidence to this perspective. First, we found evidence that the ERS latent class factor scores are significantly and fairly strongly correlated with an external 'contentless' measure of extreme response behaviour proposed by Greenleaf (1992). Second, we were able to link ERS-scores to a personality syndrome. We found that membership to a latent cluster of respondents that are sociable, extraverted, value strong opinions and view themselves as intelligent/intellectual, revealed a positive association with ERS. Also, older respondents turned out to be more inclined to make use of ERS than younger respondents.

Adding up the evidence, i.e. (a) the fact that ERS shows up regardless of length of the response scale; (b) the fact that ERS scores correlate with external measures of extreme response behaviour; (c) the fact that ERS scores are logically associated with a specific personality profile; and (d) the fact that ERS scores are associated with age; we may conclude 'beyond reasonable doubt' that ERS is much more a *response style* than a *response set*.

An important secondary goal to our research was the opportunity to introduce the latent class factor analysis approach as a flexible tool in diagnosing ERS and ARS simultaneously. Also, the opportunity given to implement the split-ballot experiment in a sizeable random

household panel and to use additional datasets from previous waves were strengths of this research.

Inevitably, there were also some shortcomings like the fact that ARS was not very well pronounced in our dataset while in other studies, like that of [Billiet and McClendon \(2000\)](#), in which similar items were included, it was. This might indicate that other test conditions than the ones tested in this research might be more significant. For example, survey length (and consequently fatigue of the respondents) might play a considerable role when it comes to ARS, since [Billiet and McClendon \(2000\)](#) used a much longer questionnaire in their study than we did. Therefore, an interesting recommendation for future research is to compare questionnaires with different lengths to check if reducing test length can prevent respondents from using ARS.

Any experiment inevitably isolates one test condition to test its effect on an experimental outcome while keeping other test conditions equal across groups. For this reason this study was not designed to draw a complete picture of what does or does not evoke response style/set behaviour. A first step is taken through showing that it is probably very difficult to avoid the occurrence of ERS by questionnaire design, at least as far as the number of response categories is concerned. With current advances in the methodology of diagnosing response bias and controlling for its effects on measurement, the question on how to prevent response bias might have become less prevalent. However, as with any disease: we do not need to be fine with being sick, simply because there is a cure. But if sickness cannot be avoided, we should be glad to have a cure. That is where we think the contribution of this research lies: it diagnosed an illness (ERS across different test conditions) which probably cannot be avoided since it seems to be a problem within the individual, but it presented a cure for it (the LCFA approach).

**Acknowledgments** In this paper use is made of data from the LISS panel of CentERdata. This research was supported by a grant from the Dutch Science Foundation NWO (No. 400-06-052). The first version of this paper has been presented at the AAPOR conference, Chicago, May 13–16, 2010.

## Appendix A

- (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 (–)

## Appendix B

Please use the rating scale below to describe how accurately each statement describes *you*.

- (1) Start conversations
- (2) Talk to a lot of different people at parties
- (3) Am interested in people
- (4) Sympathize with others' feelings
- (5) Take time out for others
- (6) Make people feel at ease
- (7) Am quick to understand things
- (8) Am full of ideas

Which values act as a guiding principle in your life and which values are less important to you.

- (9) Sincere and truthful
- (10) Responsible
- (11) Forgiving
- (12) Open-minded
- (13) Courageous
- (14) Helpful
- (15) Loving
- (16) Independent
- (17) Wisdom

Please indicate to what extent the following statements are characteristic of you.

- (18) I like to have strong opinions even when I am not personally involved
- (19) I would rather have a strong opinion than no opinion at all
- (20) I pay a lot of attention to whether things are good or bad
- (21) I want to know exactly what is good and bad about everything
- (22) I am pretty much indifferent to many important issues

Please indicate to what extent you agree or disagree with the statements below.

- (23) I feel that I have a number of good qualities

## References

- Austin, E.J., Deary, I.J., Egan, V.: Individual differences in response scale use: mixed Rasch modelling of responses to NEO-FFI items. *Pers. Individ. Differ.* **40**, 1235–1245 (2006)
- Billiet, J.B., Davidov, E.: Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociol. Methods Res.* **36**, 542–562 (2008)
- Billiet, J.B., McClelland, M.J.: Modelling acquiescence in measurement models for two balanced sets of items. *Struct. Equ. Model.* **7**, 608–628 (2000)
- Bogner, F.X., Wiseman, M.: Towards measuring adolescent environmental perception. *Eur. Psychol.* **4**, 139–151 (1999)
- Bolt, D.M., Johnson, T.R.: Addressing score bias and differential item functioning due to individual differences in response style. *Appl. Psychol. Meas.* **33**, 335–352 (2009)
- Chen, C., Lee, S.-y., Stevenson, H.W.: Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychol. Sci.* **6**, 170–175 (1995)
- Cheung, G.W., Rensvold, R.B.: Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *J. Cross-Cult. Psychol.* **31**, 187–212 (2000)

- Couch, A., Keniston, K.: Yeasayers and naysayers: agreeing response set as a personality variable. *J. Abnorm. Soc. Psychol.* **60**, 151–174 (1960)
- De Jong, M.G., Steenkamp, J.-B.E.M., Fox, J.-P., Baumgartner, H.: Using item response theory to measure extreme response style in marketing research: a global investigation. *J. Mark. Res.* **45**, 104–115 (2008)
- Diamantopoulos, A., Reynolds, N.L., Simintiras, A.C.: The impact of response styles on the stability of cross-national comparisons. *J. Bus. Res.* **59**, 925–935 (2006)
- DiStefano, C., Motl, R.W.: Personality correlates of method effects due to negatively worded items on the Rosenberg self-esteem scale. *Pers. Individ. Differ.* **46**, 309–313 (2009)
- Dolcinar, S., Grün, B.: Analytical robustness in cross-cultural comparisons. *Int. J. Cult. Tour. Hosp. Res.* **1**, 140–160 (2007)
- Dolcinar, S., Grün, B.: Response style contamination of student evaluation data. *J. Mark. Educ.* **31**, 160–172 (2009)
- Gibbons, J.L., Zellner, J.A., Rudek, D.J.: Effects of language and meaningfulness on the use of extreme response style by Spanish-English bilinguals. *Cross-Cult. Res.* **33**, 369–381 (1999)
- Greenleaf, E.A.: Measuring extreme response style. *Public Opin. Q.* **56**, 328–351 (1992)
- Hamilton, D.L.: Personality attributes associated with extreme response style. *Psychol. Bull.* **69**, 192–203 (1968)
- Harzing, A.-W.: Response styles in cross-national survey research: a 26-country study. *Int. J. Cross-Cult. Manag.* **6**, 243–266 (2006)
- Heide, M., Grønhaug, K.: The impact of response styles in surveys: a simulation study. *J. Mark. Res. Soc.* **34**, 215–230 (1992)
- Hui, C.H., Triandis, H.C.: Effects of culture and response format on extreme response style. *J. Cross-Cult. Psychol.* **20**, 296–309 (1989)
- Hurley, J.R.: Timidity as a response style to psychological questionnaires. *J. Psychol.* **132**, 202–210 (1998)
- Institute of Social and Political Opinion Research (ISPO): 1995 general election study Belgium-Flanders. <http://www.data-archive.ac.uk/findingData> (1997). Accessed 1 Oct 2009
- International Social Survey Programme (ISSP): family and gender roles III. <http://www.issp.org/data.shtml> (2002). Accessed 1 Oct 2009
- Johnson, T.R., Kulesa, P., Cho, Y.I., Shavitt, S.: The relation between culture and response styles: evidence from 19 countries. *J. Cross-Cult. Psychol.* **36**, 264–277 (2005)
- Kieruj, N.D., Moors, G.: Variations in response style behavior by response scale format in attitude research. *Int. J. Public Opin. Res.* <http://ijpor.oxfordjournals.org/content/early/2010/07/23/ijpor.edq001.full.pdf+html> (2010). Accessed 31 Aug 2010
- Knowles, E.S., Nathan, K.T.: Acquiescent responding in self-reports: cognitive style or social concern? *J. Res. Pers.* **31**, 293–301 (1997)
- Koson, D., Kitchen, C., Kochen, M., Stodolosky, D.: Psychological testing by computer: effect on response bias. *Educ. Psychol. Meas.* **30**, 808–810 (1970)
- Krosnick, J.A.: Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* **5**, 213–236 (1991)
- Lewis, N., Taylor, J.: Anxiety and extreme response preferences. *Edu. Psychol. Meas.* **15**, 111–116 (1955)
- Marín, G., Gamba, R.J., Marín, B.: Extreme response style and acquiescence among Hispanics. *J. Cross-Cult. Psychol.* **23**, 498–509 (1992)
- Meisenberg, G., Williams, A.: Are acquiescent and extreme response styles related to low intelligence and education? *Pers. Individ. Differ.* **44**, 1539–1550 (2008)
- Moors, G.: 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. *Qual. Quant.* **37**, 277–302 (2003)
- Morren, M.: The survey response: a mixed method study of cross-cultural differences in responding to attitude statements (Doctoral dissertation). Tilburg University, Tilburg, The Netherlands (2011)
- Naemi, B.D.: Measuring and predicting extreme response style: a latent class approach. <http://scholarship.rice.edu/bitstream/handle/1911/17901/1435749.PDF?sequence=1> (2006). Accessed 31 Aug 2010
- Naemi, B.D., Beal, D.J., Payne, S.C.: Personality predictors of extreme response style. *J. Pers.* **77**, 261–286 (2009)
- Pedersen, D.M.: Acquiescence and social desirability response sets and some personality correlates. *Edu. Psychol. Meas.* **27**, 691–697 (1967)
- Rorer, L.G.: The great response-style myth. *Psychol. Bull.* **63**, 129–156 (1965)
- Shulman, A.: A comparison of two scales on extremity bias. *Public Opin. Q.* **37**, 407–412 (1973)
- Smith, P.B.: Acquiescent response bias as an aspect of cultural communication style. *J. Cross-Cult. Psychol.* **35**, 50–61 (2004)

- Thompson, S.C.G., Barton, M.: Ecocentric and anthropocentric attitudes toward the environment. *J. Environ. Psychol.* **14**, 149–157 (1994)
- Van Herk, H., Poortinga, Y.H., Verhallen, T.M.M.: Response styles in rating scales: evidence of method bias in data from six EU countries. *J. Cross-Cult. Psychol.* **35**, 346–360 (2004)
- Van Rosmalen, J., Van Herk, H., Groenen, P.J.F.: Identifying unknown response styles: a latent class bilinear multinomial logit model. *J. Mark. Res.* **47**, 157–172 (2007)