Response Styles in Survey Research: A Literature Review of Antecedents, Consequences and Remedies

Yves Van Vaerenbergh
Human Relations Research Group
Hogeschool-Universiteit Brussel, Brussels, Belgium
&
Center for Service Intelligence
Ghent University, Ghent, Belgium
email: yves.vanvaerenbergh@hubrussel.be

Troy D. Thomas
Department of Political and Social Sciences
Ghent University; Ghent, Belgium
&
Department of Mathematics, Physics and Statistics
University of Guyana; Georgetown, Guyana
email: troy.thomas@uog.edu.gy

Published in *International Journal of Public Opinion Research,* 25(3), 195-217

Author Note

Yves Van Vaerenbergh is a teaching and research assistant at the Human Relations Research Group at the Hogeschool-Universiteit Brussel, and member of the Center for Service Intelligence, Ghent University. His research interests include service failure and recovery encounters, service language, and response styles in survey research. His work is published or forthcoming in *Journal of Service Research*, *European Journal of Marketing*, *Service Industries Journal*, and *Advances in Advertising Research*.

Troy Thomas is a researcher at the Department of Political Science, Ghent University and he teaches in the Department of Mathematics, Physics and Statistics at the University of Guyana. His research interests include measurement errors in surveys inclusive of response styles and measurement invariance in cross-cultural research.

Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies

Abstract

Although the purpose of questionnaire items is to obtain a person's opinion on a certain matter, a respondent's registered opinion may not reflect his or her 'true' opinion because of random and systematic errors. Response styles (RS) are a respondent's tendency to respond to survey questions in certain ways regardless of the content, and they contribute to systematic error. They affect univariate and multivariate distributions of data collected by rating scales and are alternative explanations for many research results. Despite this, RS are often not controlled in research. This article provides a comprehensive summary of the types of RS, lists their potential sources, and discusses ways to diagnose and control for them. Finally, areas for further research on RS are proposed.

Keywords: response styles, literature review, antecedents, consequences, measurement

Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies

In several social sciences disciplines, questionnaire data are indispensable sources of information. Researchers rely on respondents' self-reports to understand their attitudes and behaviours. A popular way to measure these attitudes and behaviours is to use rating scales (Moors, 2010). However, after respondents have provided their ratings for given statements, the question of whether the given answers reflect their true opinions remains.

Researchers agree that a response variance can be decomposed into true and error variances (Smith, 2011), the latter of which includes variance due to response styles (RS). Thus, RS distort research results. RS are the respondent's systematic tendency to respond to a range of survey items on a different basis from what the items are designed to measure (Paulhus, 1991). RS are present in the entire data set and they affect the validity of research conclusions in two main ways (Baumgartner & Steenkamp, 2001). First, RS affect *univariate distributions*—that is, RS have an impact on means and variances. For example, previous research has typically found gender differences in passive/laissez-faire leadership. However, Moors (2012) finds that women are more likely to use the highest and the lowest response categories of a rating scale (extreme RS) than men, which introduces systematic error into the research results. Consequently, the relationship between gender and leadership styles is spurious when taking RS into account. Thus, without controlling for RS, researchers might draw incorrect conclusions from comparative tests such as *t*-tests or *F*-tests (Chueng & Rensvold, 2000). Second, RS affect *multivariate distributions*. For example, Baumgartner and Steenkamp (2001) correlate health consciousness (HCO), quality consciousness (QCO), environmental consciousness (ECO), and ethnocentrism (ETN) and find the following correlations: HCO–QCO: 0.40; HCO–ECO: 0.33; QCO–ECO: 0.31; HCO–ETN: 0.28; QCO–ETN: 0.19; and ECO–ETN: 0.15. From a theoretical perspective, one might assume that significant correlations exist among health consciousness, quality consciousness, and environmental consciousness, but not with ethnocentrism. However, controlling for RS substantially reduces the correlations to the following: HCO–QCO: 0.20; HCO–ECO: 0.15; QCO–ECO: 0.13; HCO–ETN: 0.02; QCO–ETN: 0.00; and ECO–ETN: 0.01. Thus, RS affect the magnitude of correlations between variables. Many statistical techniques, such as Cronbach's alpha, regression analysis, factor analysis, and structural equation modelling, rely on correlations between variables (Reynolds & Smith, 2010). As a result, studies examining such relationships without controlling for RS might yield misleading results.

Therefore, RS potentially affect all empirical studies that use rating scales and are alternative explanations for the results. However, despite its importance, many researchers do not control for this source of bias. The purpose of this article is to provide insights into RS by (1) defining different types of RS, (2) discussing the different sources of RS, and (3) providing an overview of various statistical remedies for RS. This is important because, to our knowledge, no comprehensive discussion of RS is available in the literature. Given that only a few research articles control for RS, this article provides the necessary background and tools for researchers to assess RS in their own research projects.

Types of Response Styles

The literature distinguishes between several types of RS. Table 1 provides an overview of eight RS that are prominent in the literature. Included are acquiescence response style (ARS), disacquiescence response style (DARS), mid-point response style (MRS), extreme response style (ERS), mild response style (MLRS), net acquiescence response style (NARS), response range (RR), and non-contingent response style (NCRS), along with short descriptions, graphical representations when applicable, an overview of

the main consequences, and sources from which further explanations can be obtained. As Table 1 indicates, RS have various influences on observed means and/or variances and on the magnitude of the relationships between variables.

[Insert Table 1 about here]

Researchers have devoted attention mainly to investigating ARS, DARS, ERS, and MRS (Cabooter, 2010; Weijters, 2006). In the remainder of this article, we focus on these four types but also elaborate on other types when necessary.

Sources of Response Styles

Weijters (2006) classifies sources of RS into two main categories: the stimulus level and the respondent level. At the stimulus level, RS are viewed as a consequence of the survey instrument. At the respondent level, RS are viewed as a consequence of personal characteristics. Baumgartner and Steenkamp (2001) note that situational factors can encourage or discourage people's inherent tendency to use RS. Therefore, although we discuss stimulus (situational) and respondent factors separately in the subsequent section, it should be kept in mind that these factors cannot be viewed as independent of each other.

Stimuli as Sources of Response Styles

According to Maxey and Sanford (1992, p. 295), "It seems almost impossible to escape the possibility that questionnaire items influence the responses given by respondents". This suggests that questionnaire design and questionnaire items themselves act as stimuli to respondents, and therefore they may also influence RS. Table 2 summarises research on stimuli as sources of RS. These stimuli include scale format, mode of data collection, cognitive load, interviewer effects, survey language, and topic involvement.

[Insert Table 2 about here]

Scale format. Greenleaf (1992a) and Baumgartner and Steenkamp (2001) suggest examining RS for different scales formats, and some researchers have responded. For example, Kieruj and Moors (2010) find that MRS emerges when nine or more response categories are offered, and Kieruj and Moors (2012) find week evidence of ARS in 5- to 11-point rating scales. Added to this, Weijters, Cabooter, and Schillewaert (2010) find that longer rating scales have no effect on NARS but that NARS increases with the addition of a neutral point and with fully labelled scales.

For ERS, the evidence is mixed. Arce-Ferrer (2006) finds no difference in ERS between one- and two-stage rating scales, while Albaum, Roster, Yu, and Rogers (2007) find higher ERS in two-stage than in one-stage rating scales. One-stage scales are simple scales, while two-stage scales have a more in-depth question following an initial filter question. Researchers examining the impact of scale format on ERS have focused mainly on one-stage rating scales. Kieruj and Moors (2010, 2012) compare 5- to 11-point rating scales and find no effect of the number of response categories on ERS, but Weijters et al. (2010) use 4- to 7-point scales and find that ERS decreases as the number of response categories increases. While Kieruj and Moors use latent-class confirmatory factor analysis (LCFA) to model ERS, Weijters et al. use representative indicators of RS (RIRS). In addition, Kieruj and Moors label only the endpoints of the scales, while Weijters et al. contrast fully labelled and endpoint-labelled scale formats and find that fully labelled scales reduce ERS. This potentially explains why Kieruj and Moors find no differences in ERS across scale formats.

According to Weijters et al. (2010), the optimal number of response categories depends on the purpose for which the scale is to be used. If a researcher wants to report direct summaries of responses, such as means or percentages, Weijters et al. (2010) suggest the use of fully labelled 5-point (or 7-point) scales because labelling makes the scale more directly interpretable. This recommendation coincides with that of Krosnick (1999), who contends that fully labelled formats maximise reliability and validity because the labels clarify the meaning of the scale. If instead the researcher wants to relate variables or

estimate linear models, Weijters et al. (2010) suggest that the endpoint-labelled 5-point (or 7-point) rating scale is best because respondents use such scales in a way that conforms better to linear models. Response scales have also been examined with many other criteria—for example, reliability, information recovery, distribution of scale means, and ease of use (see Preston & Colman, 2000; Weng, 2004)—resulting in similar recommendations with respect to the optimal number of response categories.

Modes of data collection. Differences in RS among modes of data collection lead to important implications for researchers. Telephone surveys lead to higher ARS and ERS and lower MRS than face-to-face, paper-and-pencil, and web surveys (Jordan, Marcus, & Reeder, 1980; Weijters, Geuens, & Schillewaert, 2008). These findings suggest that the mode of data collection influences research results even when only one mode of data collection is used. If researchers use telephone surveys, they should interpret raw mean scores and variances cautiously. Mode effects on RS are also important in light of the increased popularity of mixed-mode surveys (Heerwegh, 2009). Researchers using mixed-mode data collections should be cautious about combining data coming from different modes because RS might induce observed differences in the results. Therefore, researchers should account for RS in the analysis of mixed-mode data.

Cognitive load. To our knowledge, only two studies have focused on the relationship between cognitive load and RS. Knowles and Condon (1999) find that ARS increases with cognitive load, and Cabooter (2010) finds that NARS increases with cognitive load. Cognitive load is present in many situations, and researchers should try to avoid it. Researchers can do so by inviting respondents to participate in lab research, allowing respondents to participate when they have time available, or providing a room where they can relax, to name a few (Cabooter, 2010). Researchers should also word survey questions clearly, as sub-optimal question wording requires more cognitive effort to understand the meaning of the questions (Lenzner, Kaczmirek, & Galesic, 2011). If researchers suspect that the respondents completed a survey under high cognitive load, they should conduct a post hoc assessment of RS.

Interviewer effects. Interviewer effects on RS have received limited attention in the literature. Olson and Bilgen (2011) find that experienced interviewers influence higher levels of ARS, but Hox, De Leeuw, and Kreft (1991) find no such effect. Despite the potential effect of interviewer experience on ARS, in general experienced interviewers decrease measurement errors from other sources, such as non-response (Lipps & Pollien, 2011) or social desirability (Cleary, Mechanic, & Weiss, 1981). Experienced interviewers are therefore preferred, but researchers should still control for RS.

Survey language. In general, researchers should adapt questionnaires to the local language (Usunier, 2011); however, administering questionnaires in a second language leads to lower levels of ARS and ERS but higher levels of MRS and RR than when administered in a native language (Gibbons, Zelner, & Reduk, 1999; Harzing, 2006). Overall, respondents make better use of the entire scale when responding to surveys in their native language, instead of mainly using the scales' midpoint. These findings are important because cross-cultural studies often administer questionnaires in English across different language groups (Rowland, Naidoo, Abdulkadir, Moraru, Huang, & Pau, 2010). Preferably, respondents should complete surveys in their native language because they are better able to qualify their answers on rating scales. Nevertheless, a post hoc assessment of ARS and ERS is necessary.

Topic involvement. Although topic involvement is perhaps more a task characteristic than a stimulus, we consider it because it is related to the content of the question. If an item or question is not relevant to a respondent, there will be lower involvement, which influences RS. For example, Gibbons et al. (1999) report that ERS is more prevalent if the respondent is more involved with the presented stimulus.

Respondents as Sources of Response Styles

Researchers who subscribe to the view that RS are due to the respondent argue that RS are mainly determined by the respondent's characteristics and personality. We first consider demographic variables and then explore personality and culture.

Education. With few exceptions, research indicates that education is inversely related to RS. Meisenberg and Williams (2008) find this to be nearly a worldwide phenomenon for ARS and ERS. However, research findings are not unanimous, and not all RS are investigated by each researcher. Weijters, Geuens, and Schillewaert (2010b) focus on ARS, DARS, ERS, and MRS and, except for DARS, find inverse relationships to education. However, Moors (2008) and De Jong, Steenkamp, Fox, and Baumgartner (2008) find no effect of education on ERS.

Matarazzo and Herman (1984) indicate that education is correlated with IQ and suggest that in cases of extreme absence of data, the level of education can be used as an indicator of IQ. Therefore, some link exists between education and IQ. For ERS, Light, Zax, and Gardiner (1965) find a negative relationship with IQ. In addition, they find lower MRS among older people with high IQ than younger people with high IQ but find the reverse for low-IQ people. In this case, the ages included ranged from 9 to 18 years. In addition, with intelligence measured by the American College Exam, Zuckerman and Norton (1961) find that ARS decreases as intelligence increases.

Age. Research has also questioned whether a relationship exists between age and RS (Stukovsky, Palat, & Sedlakova, 1982). For ARS, research shows evidence of a positive relationship with age (Billiet & McClendon, 2000; Greenleaf, 1992a; Ross & Mirowsky, 1984; Weijters et al., 2010b), but Eid and Rauber (2000) report no effect. The evidence for ERS is particularly interesting. Several researchers find that ERS increases with age (Greenleaf, 1992b; Meisenberg & Williams, 2008; Weijters et al., 2010b), others find that older respondents have lower levels of ERS (Austin, Deary, & Egan, 2006; Light et al., 1965), and still others find no effect (Johnson, Kulesa, Cho, & Shavitt, 2005; Moors, 2008). However, De Jong et al. (2008) find that both the younger and the elderly respondents have higher levels of ERS than the middle-aged group. This curvilinear relationship potentially explains the different findings. For example, if there is a higher proportion of elderly respondents than younger and middle-aged respondents and elderly respondents have higher ERS, one might assume a positive linear relationship between age and ERS. Conversely, if the proportion of younger respondents is higher and the younger respondents have higher ERS, a negative linear relationship with age might be assumed. Alternatively, if the proportions of younger and elderly respondents are about equal and the two groups both have higher ERS, linear modelling should find no effect. For DARS and MRS, Weijters et al. (2010b) find no effect and a positive relationship for age, respectively.

Gender. Some studies report higher ARS for women than men (Austin et al., 2006; Weijters et al., 2010b), while others report no gender effect (Light et al., 1965; Marin, Gamba, & Marin, 1992). For ERS, the results include a greater tendency among women (De Jong et al., 2008; Weijters et al., 2010b), a greater tendency among men (Harzing, 2006; Meisenberg & Williams, 2008), and no gender effect (Grimm & Church, 1999; Light et al., 1965; Marin et al., 1992; Moors, 2008). For DARS, Crandall (1973) finds no relationship with gender. For MRS, Harzing (2006) finds higher levels among women, but Light et al. (1965) and Grimm and Church (1999) find no gender effect.

Income and employment. In general, ARS and ERS are higher when socio-economic status and income are lower (Greenleaf, 1992a, 1992b; Meisenberg & Williams, 2008; Ross & Mirowsky, 1984). In addition, Johnson et al. (2005) indicate that length of employment is positively related to ARS but not to ERS. Contrary to the latter, Eid and Rauber (2000) find a positive relationship between length of employment and ERS.

Race. Prior research has found that race is a significant antecedent of RS. For example, some studies indicate that African Americans and Hispanics exhibit higher levels of ARS and ERS than White Americans (Bachman & O'Malley, 1984; Marin et al., 1992). Baron-Epel, Kaplan, Weinstein, and Green (2010) also report that ERS and MRS are higher among Jews than Arabs in Israel. These findings suggest that RS might be higher among minority groups. However, Naemi, Beal, and Payne (2009) find no support for such a conclusion regarding ERS.

In general, the literature indicates that socio-demographic variables affect RS, which suggests that researchers should be careful when comparing results across demographic profiles. However, the findings are not always consistent. A potential explanation is that empirical findings on the relationships between socio-demographic variables and RS are mere reflections of personality (Moors, 2008).

Personality. Support for the stability and consistency of RS in the literature, stability throughout the questionnaire (Hamilton, 1968, in relation to ERS), consistency throughout the questionnaire (Naemi et al., 2009; Weijters et al., 2010a, in relation to ARS and ERS), stability between data collections with a one-year time gap (Weijters et al., 2010b, in relation to ARS, DARS, MRS, and ERS), and stability over a four-year period with the same respondents (Billiet & Davidov, 2008, in relation to ARS) might be enough to counter Rorer's (1965) rejection of the notion that personality affects RS. In addition, previous research has found that ERS is positively related to intolerance of ambiguity (Brengelman, 1960; Naemi et al., 2009), preference for simple thinking and decisiveness (Naemi et al., 2009), and the Big Five personality traits extraversion and conscientiousness (Austin et al., 2006). Furthermore, Ayidiya and McClendon (1990) report a positive relationship between MRS and evasiveness, and Couch and Keniston (1960) find that ARS is positively related to impulsiveness and extraversion.

However, all previous findings on the role of personality have been criticised because rating scales are used to assess personality, and thus the personality measures themselves might be contaminated with RS (Bentler, Jackson, & Messick, 1971). Naemi et al.'s (2009) attempt to let a close friend complete the personality measures does not overcome this limitation. Conversely, Cabooter (2010) investigates 'self-regulatory focus' and ERS and MRS with the use of unique, scale-free personality measures and finds that a prevention focus is positively related to MRS and a promotion focus is positively related to ERS. These findings validate the existence of relationships between personality and RS, but because nearly all research has focused on ERS, our understanding is limited to this RS.

That personality predicts RS behaviour makes it difficult, if not impossible, for researchers to prevent respondents' use of RS (Kieruj & Moors, 2010). Therefore, researchers should diagnose and correct for RS.

Culture and country-level characteristics. Many studies highlight the relationship between RS and cultural (or cross-national) differences. Clarke (2000) finds a main effect of culture on ERS and indicates that ERS varies across countries and across sub-cultures within countries. Meisenberg and Williams (2008) report that countries with low-IQ levels show higher ERS, that countries with corrupt societies show both higher ERS and ARS, and that democracy and political freedom do not affect ARS and ERS. Van Herk, Poortinga, and Verhallen (2004) find that Mediterranean countries (Greece, Italy, and Spain) have higher ARS and ERS than Western European countries (England, Germany, and France). They also conclude that ARS and ERS increase as individualism—one of Hofstede's dimensions—decreases. However, they do not include all of Hofstede's dimensions (individualism, uncertainty avoidance, masculinity, and power distance; see Hofstede, 2001), and since the groups of countries may vary on the other dimensions, the effect of individualism is

not unequivocally established. Grimm and Church (1999) find no consistent effect of individualism on ARS or ERS and no effect of culture on MRS, while Johnson et al. (2005) find that the four dimensions are each negatively related to ARS and that power distance and masculinity are positively correlated with ERS. In addition to this, De Jong et al. (2008) find a positive relationship between ERS and individualism, uncertainty avoidance, and masculinity, while Chen et al. (1995) report a negative relationship between MRS and individualism.

Harzing (2006) examines the effect of RS on cultural variables by including both Hofstede's variables and variables based on the GLOBE dimensions (see House, Hanges, Javidan, Dorfman, & Gupta, 2004, for a description of the GLOBE dimensions). Harzing uses the GLOBE values for power distance, in-group collectivism, institutional collectivism, and uncertainty avoidance in two categories (values, or 'what should be', and practices, or 'what is'), resulting in eight variables. The findings indicate that both the nature of the relationships (whether positive or negative) and whether the relationships can be generalised (statistical significance) sometimes depend on the method of calculation (Hofstede or GLOBE values).

The relationships between culture and RS have important implications for cross-cultural (or cross-national) research. Given that obtained means, variances, and covariances are biased by RS (Baumgartner & Steenkamp, 2001), traditional measurement equivalence tests should be corrected for RS. For example, Welkenhuysen-Gybels, Billiet, and Cambré (2003) and Kankaraš and Moors (2011) demonstrate that the results of measurement equivalence tests can change substantially when adjustments are made respectively for ARS and ERS.

Overall, demographic and personality variables explain a relatively small proportion of the variance of RS, while culture and country-level characteristics seem to explain a relatively large proportion of RS in cross-cultural studies. Using a Belgian sample, Weijters et al. (2010b) find that demographic variables explain between 1.4% and 8.3% of the variance in RS depending on which RS is considered, while Meisenberg and Williams (2008) find that socio-demographic variables (for example, corruption, gross domestic product) explain approximately 1% to 5% of the variance in ARS and ERS at the individual level but that country characteristics explain approximately 63.2% (ARS) to 74.5% (ERS) at the country level. In addition, De Jong et al. (2008) indicate that Hofstede's dimensions explain approximately 59% of the between-country variance in ERS. However, because Hofstede and McCrae (2004) find significant and substantial correlations between each of Hofstede's cultural dimensions and personality (specifically, extraversion, conscientiousness, openness to experience, neuroticism, and agreeableness, as measured by the revised NEO personality inventory), overlap occurs between personality and culture. It is therefore not clear whether the indicated explanatory power for culture represents the unique effect of culture. Furthermore, although socio-demographics explain the smaller proportion of the variance in RS, they are still important determinants of RS. The effect of the personal antecedents varies from study to study, and so the explanatory power also likely varies. Neglecting socio-demographic variables as a means of controlling for RS when the data differ in relation to demographics is potentially damaging to research.

Diagnosing and Remedying Response Styles

The literature identifies several ways to diagnose and control RS. Table 3 provides an overview of the different approaches. In comparing the different techniques, several remarks are appropriate.

[Insert Table 3 about here]

First, counting double agreements on reversed items (Johnson et al., 2005), or specifying a method factor on balanced-scale items (Billiet & McClendon, 2000),

requires the use of balanced-scale items. This may be problematic, because it is often difficult to formulate reversed items (Billiet & McClendon, 2000) and because the way people respond to reversed items may be due to interpretational issues rather than ARS (Wong, Rindfleisch, & Burroughs, 2003). For example, respondents tend to minimise retrieval of additional information when answering nearby non-reversed items but tend to maximise retrieval of new and different information when answering nearby reversed items (Weijters, Geuens, & Schillewaert, 2009). As a result, balanced scales introduce several other problems that may affect the validity of the research results. Moreover, the majority of measurement scales are not balanced (Baumgartner & Steenkamp, 2001), so these techniques may not always be applicable.

Second, not all approaches simultaneously account for multiple types of RS. Multi-trait–multi-method models account for ARS and DARS but not ERS or MRS (Saris & Aalberts, 2003). The balanced-scale method (Billiet & McClendon, 2000) accounts only for ARS, while the most recently developed LCFA approach (Kieruj & Moors, 2010, 2012) allows for detection and control of ARS and ERS. The most comprehensive way to detect and control RS to date is to add representative indicators of RS (RIRSs) to the questionnaire, which allows for calculation of ARS, DARS, ERS, MRS, and NARS (Baumgartner & Steenkamp, 2001; Weijters et al., 2008). In regular studies, 5 items per response style indicator should be included, but in studies explicitly focusing on RS, 10 to 14 items per response style indicator is recommended (Weijters et al., 2008). This may not always be possible because of survey length restrictions.

Third, convergent validity between methods is not well established. De Beuckelaer, Weijters, and Rutten (2010) compare the RIRS method with the more traditional method in which survey items used for substantive purposes are also used to model ARS and ERS (count procedure). The proportion of ARS is the same for the two methods, but the correlation between the methods is low to very low. In contrast, the proportion of ERS is higher with the traditional method than the RIRS method, but the correlation between the methods is moderate to strong. Convergent validity is therefore not established between the two methods. Kieruj and Moors (2012) also examine convergent validity by correlating a latent class factor, designed to measure ERS, with a RIRS measure of ERS. The two measures of ERS are moderately correlated, thus providing preliminary evidence of convergent validity between the methods, but additional research on this issue is necessary.

To control for RS, we recommend the use of the RIRS or representative indicators response styles means and covariance structure (RIRMACS) method. These methods enable tests for various types of RS and the use of RS as covariates in subsequent analyses. Moreover, the RIRMACS method allows for evaluation of convergent and discriminant validity between the various RS. Researchers may not always have the means to include additional questions in the survey, may be working on secondary data, or may not want to assume that rating scale data are continuous. In these cases, the LCFA approach provides an alternative. It allows for separation of item content from response style and does not assume interval level data, and at least for ERS, preliminary evidence of convergent validity with the RIRS method has been established. However, given the uncertainty of convergent validity across methods, researchers should use multiple methods to account for RS and to assess the stability of their findings across the methods.

Conclusion

Although the strength of RS may vary across situational and personal sources, a careful examination of the literature suggests that RS are often a serious threat to the validity of research results. Because they affect univariate and multivariate distributions, RS are alternative explanations of most research findings. We contend that researchers should do whatever they can to control for RS, to obtain more accurate results. Doing so requires both

careful examination of the context in which the research is conducted, alongside the tools used to collect data, and the use of statistical procedures to detect and control RS. Furthermore, we provide an overview that researchers can use when evaluating the potential biasing effects of RS in their own research projects.

Although researchers have gained substantial knowledge on RS, not all the issues about this important topic have been resolved, and more work is necessary to enhance understanding of this phenomenon. Next, we provide several suggestions for further research.

Directions for Further Research

Although RS have received extensive attention, more work is necessary to extend and improve understanding of its antecedents. First, many conflicting results have emerged in the literature. Therefore, a meta-analysis that examines methodological between-study variables to provide a quantitative assessment of the different findings is necessary. For example, researchers have found differences between ad hoc measures and representative indicators as measures of RS (De Beuckelaer et al., 2010), and this potentially explains the different findings in the literature.

Second, researchers should also examine the mediating variables between antecedents and RS. Such examination would provide insights into the cognitive processes underlying the relationships between the antecedents and RS. Currently, such studies are scarce (Olson & Bilgen, 2011), and thus more work remains to be conducted in this area.

Third, the adverse impact of RS on research results has recently been demonstrated. For example, Moors (2012) shows that the previously accepted relationship between gender and leadership styles is spurious when RS are taken into account. Similar work is necessary to convince researchers about the potential consequences of not controlling for RS and to update existing theories within the various fields.

Regarding the antecedents of RS, research has focused on investigating either stimulus-related or person-related variables (Weijters, 2006). However, Baumgartner and Steenkamp (2001) note that a person-related source of RS (for example, personality) may trigger or attenuate the effects of stimulus-related sources. Research should therefore examine interaction effects among antecedents.

Because we do not yet fully understand how research designs can trigger or retard the use of RS, further research on stimulus-related antecedents would be useful. Kieruj and Moors (2012) propose that survey length might trigger ARS, but research has not yet formally examined this issue. Naemi et al. (2009) find that the amount of time a respondent spends on the questionnaire significantly influences RS, and Cabooter (2010) investigates cognitive load (as time pressure) as a situational determinant of RS. However, other situation-related variables, such as mood, fatigue, or ego depletion, may affect RS, but these relationships have not been tested properly to date.

Research seems to focus on certain scale formats, and thus several opportunities for further research exist. First, it might be useful to examine culture as a moderator of the scale format–RS relationship. This would lead to identification of the scale format the suffers least from RS and which would be of substantial benefit to cross-cultural (or cross-national) research. Second, researchers could examine whether adding a ‘don’t-know’ option to the survey affects RS. Third, Tourangeau, Couper, and Conrad (2007) examine the impact of scale colour in a web survey on mean responses to a rating scale. They find that for endpoint-labelled scales, when the end points are shaded in different hues compared to the same hue, responses shift toward the high end of the scale. Research should formally examine the impact of different scale colours on RS. Research might also examine how background colours of a web survey (for example, colour of banners, background colour itself) influence RS. Fourth, research could also assess differences in RS between unipolar

and bipolar scales and between other scale formats, such as numbered and unnumbered. Tourangeau et al. (2007) indicate that the effect of shading on mean responses disappears with fully labelled scales and reduces with fully numbered scales, so there might be merit in evaluating numbered and unnumbered scales in relation to RS. Preferably, researchers should examine all these issues in a factorial design to obtain a comprehensive picture of how scale format influences RS.

In relation to person-related variables, researchers should further explore the role of personality on RS using scale-free personality tests, such as the one Cabooter (2010) developed. In addition, researchers should either use personality measures that do not overlap with culture (as Harzing, 2006, attempted for extraversion) or explicitly model the joint effect of personality and culture on RS to quantify the overlap, clarify the unique effect of personality, and provide improved estimates of the explanatory power of culture for RS.

Another important area for research is RS measurement. Only a few studies have examined the convergent validity of RS measures, though various methods have been proposed in the literature (see Table 3). Research should further examine convergent validity between methods, preferably through simulations. This can lead to determination of the best (or optimal) method of detecting and/or controlling RS. In addition, research has recently proposed instructional manipulation checks to detect satisficing (Oppenheimer, Meyvis, & Davidenko, 2009). Research could thus examine the relationships between these instructional manipulation checks and RS.

Traditional measurement equivalence tests should include corrections for RS, but researchers should control for as many RS as possible at the same time. Currently, the procedures that give the widest coverage of RS are the RIRSMAC procedure (which accounts for ARS, DARS, ERS, and MRS; Weijters et al., 2008) and the LCFA procedure (which accounts for ARS and ERS; Kieruj & Moors, 2010, 2012). The RIRSMAC procedure assumes that rating scale data are at the interval level, while the LCFA approach regards the data as categorical (ordinal). To accommodate research that does not ascribe the interval assumption to rating scale data but wants to cover RS, the LCFA method may need to be extended, or some other alternative to the RIRSMAC procedure may need to be developed. Perhaps this alternative will exhibit greater convergent validity with the method of Kieruj and Moors (2010, 2012).

According to our review of the RS literature, although researchers have already devoted considerable attention to this topic, much still needs to be learned. We hope that we have inspired researchers to continue to expand on the boundaries of knowledge on RS.

References

Albaum, G., Roster, C., Yu, J. H., & Rogers, R. D. (2007). Simple rating scale formats: Exploring extreme response. *International Journal of Market Research, 49*, 633-650.

Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme-response style. *Educational and Psychological Measurement, 66*, 374-392.

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.

Ayidiya, S., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly, 54*, 229-247.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly, 48*, 491-509.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M.S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European Journal of Public Health*, *20*, 543-548.

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

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two- dimensional interpretation of acquiescence. *Psychological Bulletin, 76*, 186-204.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods and Research, 36*, 543-562.

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

Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, *71*, 814-833.

Brengelman, J. (1960). Extreme response set, drive level and abnormality in questionnaire rigidity. *Journal of Mental Science, 106*, 171-186.

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures*. Doctoral dissertation, Ghent University.

Chen, C., Lee, S. Y. & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, *6*, 170-175.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*, 187-212.

Clarke, I. (2000). Extreme response style in cross-cultural research. *Journal of Social Behavior and Personality*, *15*, 137-152.

Cleary, P. D., Mechanic, D., & Weiss, N. (1981). The effect of interviewer characteristics on responses to a mental health interview. *Journal of Health and Social Behavior*, *22*, 183-193.

Couch, A., & Keniston, K. (1960). Yea-sayers and nay-sayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, *60*, 151-172.

Crandall, J. (1973). Sex differences in extreme response style: Differences in frequency of use of extreme positive and negative ratings. *Journal of Social Psychology, 89*, 281-293.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity, 44*, 761-775.

De Jong, M. G., Steenkamp, J. B., Fox, J. P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research, 45*, 104-115.

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment, 16*, 20-30.

Gibbons, J. L., Zelner, J. A., & Reduk, D. J. (1999). Effects of language and meaningfulness on the use of extreme response styles by Spanish-English bilinguals. *Cross-Cultural Research, 33*, 369-381.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*, 176-188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly, 56,* 328-350.

Grimm, D. D., & Church, A. T. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality, 33*, 415-441.

Hamilton, D. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin, 69*, 192-203.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross-Cultural Management*, *6*, 243-266.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, *21*, 111-121.

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hofstede, G., & McCrae, R. R (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38*, 52-88.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004) *Culture, leadership, and organizations. The Globe study of 62 Societies*. Thousand Oaks, CA: Sage.

Hox, J. J., De Leeuw, E., & Kreft, I. G. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.). *Measurement errors in surveys* (pp. 439-461). New York: Wiley.

Hurley, J.R. (1998). Timidity as a response style to psychological questionnaires. *Journal of Psychology, 132*, 201-210.

Johnson, T., 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.

Jordan, L. A., Marcus, A. S., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly, 44*, 210-222.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across Europe: A multigroup latent-class factor approach. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 7*, 68-80.

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, *22*, 320-342.

Kieruj, N. D., & Moors, G. (2012). Response style behavior: Question format dependent or personal Style? *Quality & Quantity*, forthcoming, DOI: 10.1007/s11135-011-9511-4.

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

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.

Lenzer, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of*

Public Opinion Research, *23*, 361-373.
Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of Personality and Social Psychology, 2*, 907-909.
Lipps, O., & Pollien, A. (2011). Effects of interviewer experience on components of nonresponse in the European Social Survey. *Field Methods*, *23*, 156-172.
Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross-Cultural Psychology, 23*, 498-509.
Matarazzo, J. D., & Herman, D. O. (1984). Relationship of education and IQ in the WAIS-R standardization sample. *Journal of Consulting and Clinical Psychology, 52*, 631-634.
Maxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their use in attitude research. In N. Schwarz & S. Sudman (Eds.), *Context effects in social psychology research* (pp. 279-296). New York: Springer-Verlag.
Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44*, 1539-1550.
Moors, G. (2003). Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *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.
Moors, G. (2010). Ranking the ratings: A latent-class regression model to control for overall agreement in opinion research. *International Journal of Public Opinion Research*, *22*, 93-119.
Moors, G. (2012). The effect of response style bias on the measurement of transformational, transaction, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*, 271-298.
Naemi, B. D., Beal, D., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality, 77*, 261-282.
Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, *75*, 99-114.
Oppenheimer, D., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867-872.
Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightman (Eds.), *Measurement of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.
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.
Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, *13*, 230-243.
Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin, 63*, 129-156.
Ross, C., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, *25*, 189-197.
Rowland, M. L., Naidoo, S., Abdulkadir, R., Moraru, R., Huang, B., & Pau, A. (2010). Perceptions of intimidation and bullying in dental schools: A multi-national study. *International Dental Journal*, *60*, 106-112.

Saris, W. E., & Aalberts, C. (2003). Different explanations for correlated disturbance terms in MTMM studies. *Structural Equation Modeling, 10*, 193-213.

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

Smith, T.W. (2011). Refining the total survey error perspective. *International Journal of Public Opinion Research*, *23*, 464-484.

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

Stukovsky, R., Palat, M., & Sedlakova, A. (1982). Scoring position styles in the elderly. *Studia Psychologica, 24*, 145-154.

Tourangeau, R., Couper, M., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*, 91-112.

Usunier, J.-C. (2011). Language as a resource to assess cross-cultural equivalence in quantitative management research. *Journal of World Business*, *46*, 314-319.

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.

Van Rosmalen, J., Van Herk, H., & Groenen, P. J. F. (2010). Identifying response styles: A latent-class bilinear multinomial logit model. *Journal of Marketing Research, 47*, 157-172.

Watkins, D., & Cheung, S. (1995). Culture, gender, and response bias: Analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, *26*, 490-504.

Weijters, B. (2006). *Response styles in consumer research*. Doctoral dissertation, Ghent University.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*, 236-247.

Weijters, B., Geuens, M., & Schillewaert, N. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science, 36*, 409-422.

Weijters, B., Geuens, M., & Schillewaert, N. (2009). The proximity effect: The role of inter-item distance on reverse-item bias. *International Journal of Research in Marketing*, *26*, 2-12.

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement, 34*, 105-121.

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological Methods, 15*, 96-110.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology, 34*, 702-722.

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*, 956-972.

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

Zuckerman, M., & North, J. (1961). Response set and content factors in the California F-scale and the parental attitude research instrument. *Journal of Social Psychology, 53*, 199-210.

Table 1
Types of RS

Type	Definition	Respondent's use of a seven-point rating scale[a]	Consequences	Representative studies
Acquiescence response style (ARS)	Tendency to agree with items regardless of content, only the highest response categories are used	○○○○●●●	Inflates observed means, increases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Greenleaf (1992b)
Disacquiescence response style (DARS)	Tendency to disagree with items regardless of content, only the lowest response categories are used	●●●○○○○	Deflates observed means, increases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Stenning and Everett (1984)
Mid-point response style (MRS)	Tendency to use the middle response category of a rating scale, regardless of content	○○○●○○○	Brings observed means closer to the mid-point, deflates variance, increases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Weijters, Geuens and Schillewaert (2008)
Extreme response style (ERS)	Tendency to use the highest and lowest response categories of a rating scale	●○○○○○●	Inflates (deflates) observed means variance, decreases magnitude of multivariate relationships	Baumgartner and Steenkamp (2001); Greenleaf (1992b)
Mild response style (MLRS)	Tendency to avoid the highest and lowest response categories of a rating scale. This is the complement of ERS	○●●●●●○	Brings observed means closer to the mid-point, deflates variance, increases magnitude of multivariate relationships	Hurley (1998); Moors (2008)
Net acquiescence response style (NARS)	Tendency to show greater acquiescence than disacquiescence.	-	Inflates variance, deflates observed means if negative	Baumgartner and Steenkamp (2001); Weijters, Cabooter, and Schillewaert (2010)
Response range (RR)	Tendency to use a narrow or wide range of response categories around the mean response	-	When large: inflates variance, decreases magnitude of multivariate relationships	Greenleaf (1992b)
Noncontingent responding (NCR)	Tendency to respond to items carelessly, randomly, or non-purposefully	-	No a priori hypotheses about the effect can be specified	Baumgartner and Steenkamp (2001); Watkins and Chueng (1995)

[a] A seven-point scale is used only for illustrative purposes; RS are also present in other types of rating scales. Black dots indicate the response categories a respondent is more likely to use under a certain RS.

Table 2
Stimuli as sources of RS

Source	ARS	DARS	ERS	MRS	Representative studies
Scale format	-	-	One-stage = two-stage scale formats	-	Arce-Ferrer (2006)
	-	-	Two-stage > one-stage scale formats	-	Albaum et al. (2007)
	Weak evidence of ARS in 5-, 6-, 7-, 9-, 10-, and 11-point scales	-	No difference between 5-, 6-, 7-, 9-, 10-, and 11-point scales	-	Kieruj and Moors (2010, 2012); Moors (2008)
	Longer scales have no effect on NARS	-	Longer scales lead to lower levels of ERS	-	Weijters, Cabooter and Schillewaert (2010)
	Neutral point leads to higher levels of NARS	-	Neutral point leads to lower levels of ERS	-	
	Fully labelled scales increase ERS	-	Fully labeled scales reduce ERS	-	
Mode of data collection	Telephone > face-to-face	-	Telephone > face-to-face	-	Jordan et al. (1980)
	Telephone > paper-and-pencil and web	Paper & pencil > Web	Paper-and-pencil > web	Tel < Paper & pencil and Web	Weijters, Geuens and Schillewaert (2008)
	-	-	Web > paper-and-pencil	-	Kiesler and Sproul (1986)
	Web = face-to-face	-	-	-	Heerwegh (2009)
Cognitive load	ARS increases with cognitive load	-	-	-	Knowles and Condon (1999)
	NARS increases with cognitive load	-	-	-	Cabooter (2010)
Interviewer experience	Higher with experienced interviewers	-	-	-	Olson and Bilgen (2011)
	No interviewer effects	-	-	-	Hox, De Leeuw, and Kreft (1991)
Survey language	Native language > second language	-	Native language > second language	Second language > native language	Harzing (2006)
	-	-	-	Second language > native language	Gibbons et al. (1999)
Topic involvement	-	-	Increases with higher levels of topic involvement	-	Gibbons et al. (1999)

Table 3
Methods of detecting and correcting for RS

Measurement of RS	Description	Advantages	Disadvantages	Representative studies
Count procedure	Count the number of agreements, disagreements, extreme responses, and/or mid-point responses on substantive measures across an entire questionnaire	Easy to use, no additional indicators are necessary	Only works with heterogeneous items	Bachman and O'Malley (1984); Reynolds and Smith (2010)
Counting double agreements on reversed items	Include reversed items in the questionnaire, and count the number of double agreements on the reversed items	Easy to use, no additional indicators are necessary	Sometimes difficult to formulate reversed items, people's responses to reversed items might be due to interpretational factors	Hox, De Leeuw, and Kreft (1991); Johnson et al. (2005)
Multi-trait–multi-method models (MTMM)	The same trait is repeatedly measured by means of different methods. Observed variance can be decomposed into true variance and error variance	Easy to set up, easy to use, measures net effects of ARS and DARS, no additional indicators are necessary	Gives no indication of ERS and MRS, consistency bias and memory effects might arise due to repeated measurement, problems of identification arise often	Saris and Aalberts (2003); Saris, Satorra, and Coenders (2004)
Specify method factor in Confirmatory Factor Analysis (CFA)	Specify positive and negative loadings on content factor, specify positive loadings on a method factor	Relatively easy to specify, most researchers are familiar with CFA, no additional indicators are necessary	Does not control for DARS, MRS, or ERS; requires the use of balanced scale items; all loadings on the method factor are restricted to equality in order to identify the model	Billiet and McClendon (2000); Welkenhuysen-Gybels et al. (2003)
Latent-class regression analysis	Run a latent-class regression analysis, and assess whether a method factor emerges	No additional indicators are necessary	Specific software is necessary, researchers might be unfamiliar with latent-class analysis, sometimes hard to specify	Moors (2009); Van Rosmalen, Van Herk, and Groenen (2010)

Measurement of RS	**Description**	**Advantages**	**Disadvantages**	**Representative studies**
Latent-class confirmatory factor analysis (LCFA)	Specify two method factors, one to measure ARS, one to measure ERS	No additional indicators are necessary, recent models allow discriminating ARS and ERS	Does not account for DARS and MRS, specific software is necessary, researchers might be unfamiliar with LCFA	Moors (2003, 2012); Kieruj and Moors (2010, 2012)
Item-response theory (IRT) model	Models the probability of ticking a certain response option as a function of the underlying latent variable	Allows different items to be differentially useful for measuring ERS, relaxes the assumption that ERS measures should be uncorrelated	Only developed for ERS, requires use of Markov chain Monte Carlo procedures, which might be more difficult to implement	Bolt and Newton (2011); De Jong et al. (2008)
Representative indicators for response styles (RIRS) method	Include a number of uncorrelated, maximally heterogeneous measures in content to the survey, and calculate weighted RS indicators	Easy to calculate, allows measuring ARS, DARS, ERS, MRS, NARS, not related to content, easy to include as covariates in subsequent analyses	Additional items need to be added to the survey	Baumgartner and Steenkamp (2001); Greenleaf (1992a, 1992b); Weijters (2006)
Representative indicators response styles means and covariance structure (RIRMACS) method	Add additional, uncorrelated items to the survey, which serve as observed variables in a CFA; ARS, DARS, MRS and ERS serve as latent variables; Extends the RIRS method	Easy to use, RS indicators can be added as covariates in subsequent analyses, use of specific RS indicators allows discrimination between content and style, allows measurement of ARS, DARS, MRS, and ERS; allows testing of convergent and discriminant validity of the different RS	Additional items need to be added to the survey	Weijters et al. (2008)