

Tilburg University

Response styles in rating scales

van Herk, H.; Poortinga, Y.H.; Verhallen, T.M.M.

Published in:
Journal of Cross-Cultural Psychology

Publication date:
2004

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

Citation for published version (APA):
van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from 6 EU countries. *Journal of Cross-Cultural Psychology*, 35(3), 346-360.

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.

Version: November 27, 2003

Response Styles In Rating Scales: Evidence of Method Bias in Data from 6 EU
Countries

Hester van Herk

Ype H. Poortinga

Theo M.M. Verhallen

Hester van Herk is Assistant Professor of Marketing, Vrije Universiteit Amsterdam,

Ype H. Poortinga is Professor of Psychology at Tilburg University and at Leuven

University Belgium, Theo M. M. Verhallen is Professor of Marketing and Marketing

Research at Tilburg University.

The authors thank Philips DAP, Amersfoort and Unilever, Vlaardingen, the Netherlands for kindly making available the data used in this article.

Response Styles In Rating Scales: Evidence of Method Bias in Data from 6 EU Countries

Abstract

In cross-cultural studies with social variables such as values or attitudes, it is often assumed that differences in scores can be compared at face value. However, response styles like acquiescence and extreme response style may affect answers, particularly on rating scales. In three sets of data from marketing studies, each with representative samples from at least three out of six countries (Greece, Italy, Spain, France Germany and the UK), these two response styles were found to be more present in the Mediterranean than in North Western Europe. Evidence for response style effects was not only found in response distributions on rating scales, but also in discrepancies of these distributions with national consumer statistics and self-reported actual behavior.

Key words: method bias, response styles, acquiescence, extreme response style, countries in the European Union

Running head: Response styles in EU countries

Response Styles In Rating Scales: Evidence of Method Bias in Data from 6 EU Countries

In this article we examine effects of one type of bias, namely response styles, that can affect answers on questionnaires. In cross-cultural studies, particularly with social variables such as value dimensions, one often finds that scores are being compared at face value. Such analyses assume that there is no systematic bias in these data. Various authors have alluded to the possibility of systematic differences in response styles between countries, or other populations defined in terms of culture (e.g., Berry, Poortinga, Segall & Dasen, 2002; Hui & Triandis, 1989; Johnson & Van de Vijver, 2003; Van Hemert, Van de Vijver, Poortinga & Georgas, 2002). However, little is known about the extent to which such artifacts distort the validity of cross-cultural comparisons. One reason is that analysis of response styles requires an extensive research effort, because responses have to be checked somehow against a common standard, preferably actual behavior.

In order to establish possible effects of response styles, we reanalyzed data from three multinational surveys, each including various countries of the European Union (EU), and together covering more than 6500 respondents. The surveys pertain to three behavioral domains (cooking, washing, and personal care), and include items at various levels of abstraction. In addition, consumer statistics on actual behavior were available to differentiate style effects from valid differences on the target variables.

Response bias is 'a systematic tendency to respond to a range of questionnaire items on some other basis than the specific item content' (Paulhus, 1991, p. 17). For example, a respondent may tend towards answering on the positive side of a rating scale when assessing items or choose the most socially desirable answers. If a respondent displays

bias consistently across items and methods this bias is called a response style. There are three prominent response biases in the psychological literature (Paulhus, 1991), namely socially desirable responding, acquiescence, and extreme response bias. Social desirability is the tendency to make oneself look good in terms of prevailing cultural norms when responding to questionnaire items (Mick, 1996); it is especially important in personality scales or self-reports of sensitive behavior, and is the most frequently studied response style (Paulhus, 1991). Acquiescence is the tendency to agree rather than disagree with items, regardless of item content. It is also called agreement tendency or yea-saying. Extreme response bias (Paulhus, 1991) is the tendency to endorse extreme response categories on a rating scale (e.g., the 1 and/or 5 on a five-point scale) regardless of content. In other publications, this is also called extreme response style (e.g., Crandall, 1982; Greenleaf, 1992b; Hamilton, 1968).

Response bias has been an issue in educational and psychological measurement for half a century (see Messick, 1991). Before that time, it was usually taken for granted that the answers respondents gave were a true reflection of their knowledge or their opinion. In the 1940s, research was started on the problem of response bias and Cronbach (1946, 1950) published two reviews on the issue. Cronbach (1950) used the term 'response set' to refer to tendencies in responses separately from content. The term 'set' referred to a temporary reaction or a situation demand on a respondent, such as time pressure, or a specific item format, suggesting that by employing another item format or by doing the study at another time, unbiased measures can be obtained. However, some response sets appeared to be relatively stable (Cronbach, 1946; Messick, 1968) and it was suggested that these response sets might reflect aspects of personality as well. To emphasize this point, Jackson and Messick (1958) proposed the term 'response styles' as they considered the respondent to display bias consistently

across time and situations. The terms 'response set' and 'response style' are both still used in literature; some researchers use the term 'set' (e.g., Cheung & Rensvold, 2000; Chinese Culture Connection, 1987; Forsman, 1993; Hui & Triandis, 1985), whereas others use the term 'style' (e.g., Bachman & O'Malley, 1984; Baumgartner & Steenkamp, 2001; Greenleaf, 1992a; Messick, 1991). Here we adopt the term 'response style'. With this we refer to a response tendency of an individual that has some consistency across items, across methods, and across behavior domains.

The present study addresses acquiescence and extreme response style. In attitude and survey research these are considered the most problematic (Bentler, Jackson & Messick, 1971; Schuman & Presser, 1996).

The issue of response styles is directly related to method bias as elaborated by Van de Vijver and Leung (1997) for cross-cultural psychology. They define three major kinds of bias: construct bias, method bias and item bias. The three levels of bias are related to three levels of equivalence (i.e., absence of bias). These three levels are: (i) structural or functional equivalence (a test measures the same trait cross-culturally, but not necessarily on the same quantitative scale); (ii) metric or measurement unit equivalence (measurement units of the scales are the same in all cultures, but there is no common scale anchor or origin); and (iii) scale equivalence or full score comparability (scores of a given value have in all respects the same meaning cross-culturally and can be interpreted in the same way).

When an instrument measures the same construct across cultures, the scores are said to be structurally equivalent or to have no construct bias. This is a minimum requirement for any kind of comparison. The joint properties of the scoring scale are limited to qualitative aspects; the scale pertains cross-culturally to the same domain or construct, but units of measurements (steps on the test score scale, e.g., a Likert scale)

may be different across cultures. In other words, if an instrument is structurally equivalent for two cultures A and B, and if it measures a given construct in culture A, then it will also measure that same construct in culture B. However, numerically equal scores may well reflect higher (or lower) levels of intensity in the one culture than in the other culture.

A lack of metric equivalence is likely due to method bias. This type of bias occurs when most or all items are affected to a similar extent by a factor that is independent of the construct studied. For example, if there is a general tendency towards acquiescence in some culture, the responses on all items will be affected. Method bias differs from item bias. The latter occurs when one or a few items deviate from the expectations about the response patterns in two cultural groups derived from other items in the same instrument. In other words, item bias can often be detected within a data set, for example, with the help of analyses based on item response theory (see e.g., Ellis & Kimmel, 1992). Removal of biased items should result in more equivalent scores. Method bias is not influenced by such deletions: inter-group differences in mean scores that reflect method bias will still be present.

One approach to detecting method bias consists in measuring a set of constructs (or traits) with various methods (so called multitrait-multimethod design, e.g., Bagozzi & Yi, 1991; Campbell & Fiske, 1959; Marsh & Byrne, 1993). If the methods provide the same outcomes for each trait, then systematic sources of bias are less likely. Another way to assess response style is through the use of balanced scales. Such scales contain item pairs that are logical opposites (e.g., Knowles & Nathan, 1997; Ray, 1983; Ten Berge, 1999). However, developing balanced scales is difficult (Schuman & Presser, 1996). For some items, there is no logical opposite. In addition, questionnaires in

(applied) survey research usually cover various topics, and due to financial and time constraints, the number of questions that can be asked per topic is limited.

Studies focusing on response styles in different cultural samples are scarce, and almost exclusively have employed students. For instance, response styles have been explored in Afro-American and European-American high school seniors (Bachman & O'Malley, 1984), and in groups of Hispanic and non-Hispanic Americans (Hui & Triandis, 1989; Marín, Gamba & Marín, 1992). In these studies the European-American respondents tended to display acquiescence less often. Watkins and Cheung (1995) reported less acquiescence for children in Australia compared to children in China, Nepal, and the Philippines. Grimm and Church (1999) found that Philippine students acquiesced more than American students when 2-, 5-, and 6-point rating scales were used. With samples in Europe, Williams (1991) found a higher mean in Italian samples. Steenkamp and Baumgartner (1998) reported differences in scale means between Greek, British and Belgian respondents. They noted that Greek respondents had the highest mean, followed by British and Belgian respondents. In another study, Baumgartner and Steenkamp (2001) argued that Greek and Portuguese respondents displayed more acquiescence than respondents from other EU countries.

In our opinion there is no clear theoretical account for these findings. One suggestion is that respondents in countries commonly referred to as more collectivistic (e.g., China, Philippines, Greece) tend to acquiesce more than respondents in individualistic countries¹. However, considering that acquiescence is a response set not contingent on item content, a relationship between acquiescence and the individualism-collectivism dimension may point to the latter being a style dimension rather than a value dimension (i.e., the way in which the individualism-collectivism syndrome

¹ This idea is supported by Smith (in press). Re-analyzing published studies including 34 or more nations, he found that acquiescence was positively related to collectivism.

usually is interpreted). Although it cannot be ruled out that values are related to styles (see Smith, in press), we are hesitant to postulate such broad relationships without more direct evidence of a causal link (see Berry et al., 2002).

Differences in extreme response style were found between student samples from the USA, Canada, Japan, and Taiwan by Chen, Lee and Stevenson (1995). In their study, American students had more extreme scores than Japanese and Taiwanese students, and typically avoided the midpoint of the rating scales. In other studies employing Korean and US student samples (Chun, Campbell & Yoo, 1974; Lee & Green, 1991), a similar result was found; American students displayed more extreme scoring than Koreans did. Bachman and O'Malley (1984) reported more extreme response style for Black than for White high school seniors. Comparing extreme responses of students in Egypt, Syria, and Jordan, Soueif (1968) found that Egyptian students gave more extreme responses than Jordanian and Syrian students did. Watkins and Cheung (1995) reported more variation in scores (e.g., more extreme responding) for Australian children compared to children from China, Nepal, Nigeria, and the Philippines. However, Stening and Everett (1984) found that Indonesian and Malaysian non-college graduates displayed more extreme scoring than American respondents. Thus, regarding extreme response style, no clear relationship with Hofstede's (2001) country scores for individualism was found.

To explore the two response styles and to test a possible relationship with individualism-collectivism, we used data from surveys in six countries in the EU, Greece, France, Spain, Italy, Germany, and the United Kingdom. The surveys covered three behavioral domains, cooking, washing, and personal care. We used the survey data to explore acquiescence and extreme response style across items at various levels of abstraction, including product specific attitudes, domain related attitudes, and general values. Moreover, we considered the results in the light of national statistics on actual levels of consumption and on self-reports of actual behavior.

Method

Survey I

The first survey was part of an international study on cooking behavior that was carried out in the period of 1988 through 1990 in three European countries, Greece, Italy, and France. In each country, a national quota sample of female homemakers was drawn that was representative with respect to age, region, and rural/urban residence. The sample sizes were 580, 598, and 532 in Greece, Italy, and France, respectively. Mean ages were 38.1 in Greece, 40.1 in Italy, and 41.5 in France.

Using a face-to-face data collection procedure, information was gathered on attitudes regarding cooking behavior, actual use of products, and some demographic variables. Bilinguals translated the questionnaires from English into Greek, Italian, and French. In a second step, a discussion between project coordinators of the research agencies in the participating countries was held to check the equivalence of the translations. There were 12 items on domain related attitudes and 12 items on product specific attitudes. Examples of the domain related items are 'I choose products that save me cooking time'; 'One always has to think about health when preparing a meal'; and 'I consume more calories than I need'. Five-point rating scales were used with symmetrical response categories ranging from 'strongly disagree' (= 1) to 'strongly agree' (= 5). Scale points were labeled verbally for the respondents.

The product-specific attitudes were concrete; they referred directly to a specific product (butter or margarine) used in a specific application (making pastry, spreading on bread, spreading on toast). These applications can be considered functionally equivalent in an international context at least for European countries (cf. Craig & Douglas, 2000). In total, six items per application-product combination were presented

to measure product specific attitudes. Examples of items were 'it is a product that offers value for money', and 'the product is easy to spread'. The items were measured on nine-point rating scales with symmetrical response intervals ranging from 'totally disagree' (=1) to 'totally agree' (=9) around a midpoint. Only the end points and the midpoint ('neither agree nor disagree') were labeled for the respondents. All items were formulated positively.

Survey II

This survey, carried out in 1993, was on 'washing of clothes' in six EU countries, Greece, Italy, France, Germany, the UK, and Spain. The sample sizes (in the same order) were 281, 299, 313, 295, 298, and 303. Respondents in all samples were female homemakers, aged 20 to 60. Each quota sample was representative with respect to age and family size in each country. Mean ages were 40.5 in Greece; 44.3 in Italy; 39.7 in France; 40.1 in Germany; 40.1 in the UK; and 41.8 in Spain.

Information was gathered on washing clothes, using a face-to-face data collection procedure. The respondents assessed five items on washing. All items were formulated positively. An example of an item is 'Doing the laundry for your family gives you a lot of satisfaction'. These items were assessed on a 5-point rating scale, where 1='do not agree at all' and 5='agree completely'. All scale points were labeled. The same procedures for translation were used as in Survey I.

Survey III

The third survey was held in 1996 in five EU countries, namely the UK, France, Germany, Italy, and Spain. Respondents in this survey were male members of a large

European marketing research panel. In each country, the samples were representative with respect to age. To make the age range similar to that of the two previous surveys, only men aged 20 to 60 were included in the present analysis. Adjusted final sample sizes were 580, 550, 634, 806, and 453 in the UK, France, Germany, Italy, and Spain, respectively. Mean ages were 39.9 in Germany, 39.8 in the UK, 37.5 in Spain, 39.2 in Italy, and 37.9 in France.

In this survey, a mail questionnaire was employed to collect information on attitudes related to shaving, actual shaving practices, general values, and socio-demographic variables. The study included 41 items on attitudes towards shaving. Each respondent assessed the items with his own shaving method in mind, namely electric or blade shaving. All items were formulated positively, and answered on five-point rating scales, with the endpoints labeled 1 ('disagree strongly') to 5 ('agree strongly'). Examples of items are 'You feel fresh during shaving'; and 'It is simple and easy to use'. The items cover several topics in shaving experience, such as the results and the convenience of the shaving method.

The List of Values (LOV) by Kahle (1983) was used to assess general values. The LOV is composed of nine items, which were answered on nine-point rating scales. The scales were labeled 1 ('very important') to 9 ('very unimportant'); midpoints of the rating scales were not labeled. Respondents were given the LOV items and the instruction that this 'is a list of things that some people look for or want out of life'. They were asked to study the list and then rate each item on the nine-point scale. An example of a LOV item is 'Security'.

In addition, information was asked in 10 items about actual behavior such as, for example, shaving frequency, and use of beauty care products.

In each domain covered by one of the surveys information was available on actual behavior in the various countries either from survey questions or from national statistics. For national consumer statistics, we have employed the 'Consumer Europe' database (Euromonitor, 1994, 1997); this is an important source of data on international consumer markets (cf. Craig & Douglas, 2000). The database is compiled by Euromonitor Ltd. from many hundreds of different sources, including data from national statistical offices, trade associations, and company research. Data typically refer to actual retail sales to the general public in a country in a given year. Data cover sixteen major consumer goods categories. The first edition was published in 1976.

Results

In cross-cultural assessment, a first step is to determine whether scales are actually measuring the same concepts in all countries. Specifically, construct equivalence or absence of construct bias has to be established. For this purpose we analyzed the similarity of principal components structures in the various countries. After a target rotation of this structure in each country towards the components structure across all countries, Tucker's coefficient of proportionality, called Tucker's Phi (Tucker, 1951) was calculated. This procedure for assessing factorial agreement is common in cross-cultural research. Usually .90 is taken as a lower bound for equivalence (Van de Vijver & Leung, 1997).

For the attitudes in the survey on cooking, three components were found in each country; they all had a value for Tucker's Phi .92 or higher. For the product specific attitudes on baking there were three components; all coefficients were higher than .90. For the dataset on washing with two principal components the proportionality coefficients were .97 or higher in all countries. For the items on personal care five principal components were found for both wet shaving and dry shaving. For wet

shaving all coefficients were .90 or higher. For electric shaving Tucker's Phi was below .90 in 4 countries for one component (4 items on absence of irritation), so this pointed to the presence of some construct bias. For the present analysis this component was removed from the data set. The Kahle values scale had two components, the first showed proportionality coefficients above .95 for all countries. The second component, containing the items 'Excitement' and 'Sense of Belonging' showed incongruities and was also removed from the data set.

Acquiescence and extreme response indices

To obtain scores for acquiescence and extreme response behavior, the frequencies of responses in the various response categories of a scale were calculated. Following the procedure of Bachman and O'Malley (1984), we computed the index for extreme response style as the relative number of scores given on the extreme categories of a rating scale. We counted the responses in categories 1 and 5 on the 5-point rating scales and the categories 1 and 9 on the 9-point scales. This number was divided by the number of items, resulting in an extreme response index ranging from 0.00 to 1.00.

To calculate an index for acquiescence we counted the number of clearly positive scores (2 highest categories on the rating scales) and subtracted the number of clearly negative scores (2 lowest categories on the rating scales). Thus, on the 5-point rating scales, the values 1, 2, 4, and 5 were taken and on the 9-point rating scales the values 1, 2, 8, and 9. The resulting number was divided by the total number of items, resulting in an acquiescence index ranging from -1.00 to 1.00. An advantage of the inclusion of two response categories at each end of a scale rather than one, was a much lower correlation between acquiescence index and extreme response index. The average correlation between the two indices was .23 when from both ends of a scale two categories were

included in the acquiescence index, and .45 when only one category was taken. It may be noted that correlations between measures of acquiescence and extreme response style are to be expected and commonly found (e.g., Baumgartner & Steenkamp, 2001).

The acquiescence and the extreme response index were calculated separately for each item set and for each country. For example, for respondents in each country in survey I (cooking domain) six scores were obtained, three for the domain related attitudes and three for the product specific attitudes. The reliabilities (Cronbachs α)² for the acquiescence and the extreme response index varied from .45 to .95 (Median: .79), with two exceptions. Both outliers were found in the acquiescence index; $\alpha = .20$ was found for attitudes towards washing in Greece, and $\alpha = .34$ was found for domain specific attitudes for cooking in Greece. On inspection of the scores the exceptions were clearly due to ceiling effects. No variables were removed as differences between respondents were of no interest in this study, and consistently high scores in a country resulting in ceiling effects do not affect the stability (and validity) of the country mean. The average correlation between acquiescence indices was .35 ($p < .001$), and between extreme response indices .48 ($p < .001$), pointing to a reasonable level of convergent validity.

Acquiescence

In Table 1, an overview is given of the values of the acquiescence index for the various item sets in each country. The mean scores for the index range from -.523

² The reliability of the extreme response index was calculated as follows. First, we defined new variables for each of the attitudinal statements. These variables were coded '1' if a person scored an extreme response, such as a '1' or a '5' on an original five-point scale, and were coded '0' otherwise. Second, we calculated the internal consistency of the scores per country and per domain. For calculating the reliability of acquiescence scores the original variables were coded: '1' if a person scored '4' or '5' on the original five-point scale; '-1' if a person scored '1' or '2'; and '0' otherwise. The second step is similar to the one for calculating the extreme response index. For the nine-point scales the same procedure was followed using the extremes 1, 2 and 8, 9.

(Germany, electric shaving, emotional aspects) to .951 (Greece, cooking, product specific attitudes). A negative index means that respondents more disagree than agree with the items.

For each of the eighteen item sets, an Analysis of Variance was performed to determine whether acquiescence differed across countries. Following this analysis, differences between separate countries were tested using the Tukey-HSD test. In 17 out of 18 item sets, there were significant differences between countries (F-test, all $p < .001$).

The acquiescence indices in Table 1 indicate that Greek respondents tended to answer more positively than Italian and French respondents. The Table shows substantial differences between Greece and Italy and between Greece and France on attitudes towards cooking and washing. For each of the eight item sets, the acquiescence index is higher in the Greek sample; this in itself already differs significantly from a chance distribution ($p < .01$). For Greek and French respondents the means are also different on all occasions. Similarly, differences in acquiescence indices were found between French and Italian respondents. In eleven of the eighteen item sets, the Italian and French indices are significantly different. In all eleven cases, the Italian index is higher ($p < .01$). The Spanish acquiescence index is higher than the French respondents' index in six out of six different item sets ($p < .05$). The number of times the acquiescence index in Germany and the UK is lower than in Italy and Spain consistently differs from a chance distribution ($p < .05$). The acquiescence indices for German, French, and British respondents do not show consistent differences.

The differences in acquiescence indices are only meaningful if they account for a substantive proportion of the variance (Cohen, 1988). As can be seen in Table 1, the average η^2 in this study is .074, ranging from .00 (item set on convenience in electric

shaving) to .30 (item set on product specific attitudes). Peterson, Albaum and Beltramini (1985) reported that in most studies of consumer behavior, η^2 attributable to treatments in experimental consumer research are below .090. Thus, the amount of variance accounted for by acquiescence in this study is of the same order of size as expected treatment effects.

Extreme response

Results for the extreme response index are presented in Table 2. As for acquiescence, there are significant differences between countries (F-test, $p < .001$). Results for the Greek respondents are striking. In all eight item sets in which Greek respondents participated, the extreme response index is higher for Greek than for Italian and French respondents (Tukey HSD, $p < .05$). In Italy, the extreme response index is higher than in France. In 12 out of the 13 item sets where scores are significantly different, the index is higher for Italy ($p < .01$). A similar result emerges between Italy and Germany (9 out of 9; $p < .01$) and between Italy and the UK (8 out of 8; $p < .01$). Italian respondents consistently tend to have a higher extreme response index than respondents in the Western European countries. As in Italy, the extreme response index in Spain is consistently higher than in France, Germany, and the UK ($p < .05$). Between Italy and Spain and also between France, Germany, and the UK there are no consistent differences in the Table.

The index of effect size η^2 (Cohen, 1988) is reported in the last column of Table 2. The average effect size is .071, ranging from .01 (items on emotions in electric shaving) to .28 (items on attitudes on cooking). The latter effect size is almost of 'medium' size (Cohen, 1988).

Response profiles of items

If there is no acquiescence or extreme response style, items with the same mean across countries have about the same expected distribution of responses over categories. Thus, the mean of scores endorsed in the middle of the scale ('2', '3' or '4') can be employed to predict the number of responses in category '5'. This distribution should be independent of country.

Across the three surveys, 85 items were measured on 5-point rating scales. For these items intervals of means for scores in the three middle categories were chosen in such a way that about 5 items in each country fell within an interval. Intervals that satisfied this criterion were 2.89 through 2.97, 3.02 through 3.09, 3.24 through 3.29, and 3.41 through 3.45. As to be expected the proportion³ of '5'-scores increased monotonically with the interval value. An ANOVA revealed that both interval ($F_{(3,130)} = 72.26$, $p < .001$; $\eta^2 = .56$) and country ($F_{(5,130)} = 14.04$, $p < .001$; $\eta^2 = .38$) were important main effects. The interaction was not significant. The main effect of country indicates that there were differences in response style across countries. The proportion of responses in the highest category of the scale ('5') was significantly higher for the Greek respondents for item clusters with a mean in the same interval than in other countries. Over all countries the average proportion of '5' across the four intervals was 23%. For Greek respondents, this proportion was 33%, i.e. 10% higher. For German respondents, the proportion '5' was 7% lower than the average. For the other countries, these percentages are 8% higher for Italian respondents, 2% higher for Spanish, 5% lower for French, and 5% lower for British respondents. A similar result was obtained for the items employing nine-point scales, where only one subset of five items with a

³ Proportions were taken instead of numbers of observations, because frequencies differed across surveys.

similar mean on the midpoints could be identified. The proportion of responses in category '9' was highest in Greece. All in all, these results confirm the other findings and indicate that these cannot be ascribed to a few items only; rather there are systematic differences in response styles between EU countries.

Measures of actual behavior

In order to further examine whether the observed differences reflect bias or differences in the target behaviors between countries, we used information from external sources on actual consumer behavior such as detergent sales, and sales of razors (Euromonitor, 1994, 1997). Moreover, the surveys asked also for information on practices, including self-reports on the applications used while cooking (survey I), and shaving frequency (survey III). If the observed differences in attitude scores are not due to response effects, they should match the data on actual behavior, because attitude – behavior relationships are expected to be neutral or positive (Fazio, 1986).

To gain some insight into these relationships, we calculated a Spearman rank order correlation between the response style indices and the measures of actual behavior from surveys II and III. For survey II, the correlation between measures of actual behavior and response tendencies across the 6 countries is .33 (ns) for acquiescence and .02 (ns) for extreme response style. For survey III, blade shaving, the correlation was -.50 (ns) for acquiescence and -.60 (ns) for extreme response style; for electric shaving the correlation was -.70 (ns) for acquiescence and -.90 ($p < .10$) for extreme response style. These results show that the national differences in response tendencies in ratings do not translate into national differences in actual behaviors; some of the correlations between behavior and response tendencies even are negative.

Another result making response styles (bias) a plausible explanation was found for Survey I. Here a low positive correlation was found between numbers of dishes prepared in cooking (behavior) and liking to cook (attitude) in all countries studied (r is about .12 in each country, $p < .01$). As can be seen in Table 3, the variety in cooking is highest in France, followed by Italy and Greece. Also, the number of times homemakers prepare pastry is higher in France and Italy than in Greece. Thus, it appears that the behavior of Greek homemakers does not justify a more positive attitude than the behavior of the French and Italian homemakers. In washing Greek homemakers use less detergent and other washing products than homemakers in Italy and France (Euromonitor, 1994). Again Greek respondents' attitudes do not appear to be in line with actual behavior.

Regarding shaving, the number of times per week a man shaves himself is positively correlated with his satisfaction about shaving ($r = .10$, $p < .01$). This result was consistent across all countries and both shaving methods studied. Thus, men who shave more should have (slightly) more positive attitudes towards shaving. However, behavior data (see Table 3) show that men in e.g., France shave more often than men in Italy do, although the pattern of means for attitudes was reversed. Moreover, French men spend more on razors and blades, and on electric shavers. Thus, differences in scores do not translate into corresponding differences in behavior. All in all, these results give reason to uphold the interpretation of score differences between countries as due to response styles.

Discussion

This article examined whether there are differences in response style between six countries in the European Union. An important finding is that Greek respondents tended

to have a higher acquiescence and a higher extreme response index than respondents in the other EU countries. The scores for Greek respondents were higher, independent of rating scale type and behavioral domain. This higher acquiescence tendency was also demonstrated when items with a similar mean on the middle categories of rating scales were analyzed for the proportion of extreme positive ratings. This proportion was consistently the highest in Greece. In addition, we found that differences in score distributions on the self-report attitudes measures were not consistent with actual differences in behavior found between the countries, both in consumer statistics and self-reported behavior.

Although contrasts were less strong than for the Greek samples, Spanish and Italian respondents also had consistently higher scores on acquiescence and extreme response indices than the British, German, and French samples. These latter three samples had about the same mean scores both on the acquiescence and on the extreme response index; if anything, British tended to display the lowest acquiescence of all EU respondents. Previously reported differences between Mediterranean and North Western European countries were found again in the present study. Moreover, our data suggest that such differences are likely to be indeed a matter of response styles, because in several tests no relationship was found for any relationship between higher levels of endorsement and actual behavior.

Thus, it appears that Mediterranean countries that are usually classified as more collectivistic on values dimensions like individualism-collectivism (Greece, Spain) attain higher scores on acquiescence and extremity response styles than the North Western European countries that are said to be more individualistic. If a comparison is made with Hofstede's results on individualism (1980, 2001), only Italy is somewhat of an exception. However, the individualism score found for this country by Hofstede was

substantially higher than predicted on the basis of external indicators (see Hofstede, 2001, p. 215).

We are reluctant to interpret our findings in terms of the individualism-collectivism dimension as countries were not selected to represent this dimension. Moreover, there would be two competing interpretations. The first is that acquiescence is part of a broad syndrome of individualism-collectivism, together with many other aspects of behavior (e.g., Smith, in press; Triandis, 1989, 1995). The other interpretation is that the many findings mentioned in the cross-cultural literature as supporting the individualism-collectivism distinction have come about, at least in part, through cross-cultural differences in acquiescence, and perhaps other response styles, supporting suggestions made, for example, by Berry et al. (2002) and by Van Hemert et al. (2002). Needless to say that the latter interpretation affects the validity of the distinction as a value syndrome.

The focus in our analysis was on acquiescence and extreme response style. One could argue that the observed differences could at least in part be due to social desirability effects. As mentioned in the introduction, this response style refers to a tendency to enhance one's self-image, i.e., to make oneself look good in the eyes of others, for example the interviewer. Items on 'healthy cooking' may have been affected by social desirability. However, for items on product characteristics of butter or margarine an explanation in terms of social desirability appears to be less plausible. Agreeing that butter or margarine is 'easy to spread' does not seem to make the respondent look good. Another reason making social desirability effects less likely can be found in data from Italy and France. Acquiescence is about equally present in the surveys (cooking and washing) where face-to-face interviewing was used, and in the survey where a mail questionnaire was employed. If social desirability would be a major explanation, one

might expect that the effect should have been stronger in face-to-face data collection. All in all, we think that the present findings have both theoretical and practical implications. Theoretically they demonstrate that method bias can play a significant role in cross-cultural research, even with countries that by and large have a fairly similar standing on major dimensions, like economic affluence, education and political organization. We think that our findings are extremely worrisome for the many cross-cultural studies in which observed differences on rating scales are interpreted at face value.

Practical implications are evident in the field of marketing from where the data were obtained. In analyses to predict sales volumes, it is common practice to employ score levels of surveys as an indication of purchase intentions. All other factors being equal, a company should not expect the same sales in Southern European countries as in North Western European countries when equal mean scores on attitude scales are being found.

In summary, our findings indicate that there are acquiescence and extreme response effects in EU countries. Our results showed consistency across rating scales and more often than not the results were consistent across the three behavioral domains investigated. Perhaps the most important contribution of this article is the demonstration that rating scale scores did not match differences in actual behavior between countries. This strongly confirms that ignoring national differences in response styles may lead to invalid inferences in cross-cultural research.

References

- 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.
- Bagozzi, R.P & Yi Y. (1991). Multitrait-multimethod matrices in consumer research. *Journal of Consumer Research*, 17(March), 426-439.
- Baumgartner, H., & Steenkamp J.B.E.M. (2001). Response styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2), 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(3), 186-204.
- Berry, J.W., Poortinga, Y.H., Segall, M.H., Dasen, P.R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). Cambridge: Cambridge University Press.
- Broen, W.E., & Wirt R.D. (1958). Varieties in response sets. *Journal of Consulting Psychology*, 22(3), 237-240.
- Campbell, D.T., & Fiske D.W. (1959). Convergent and discriminant validity by the multitrait - multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- 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(3), 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(2), 188-213.

Chinese Culture Connection (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology*, 18(2), 143-164.

Chun, K-T, Campbell J.B., & Yoo J.H. (1974). Extreme response style in cross-cultural research. A reminder. *Journal of Cross-Cultural Psychology*, 5(4), 465-480.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale New Jersey: Lawrence Erlbaum Associates.

Craig C.S., & Douglas S.P. (2000). *International Marketing Research*. Second Edition, New York: Wiley.

Crandall, J.E. (1982). Social interest, extreme response style, and implications for adjustment, *Journal of Research in Personality*, 16, 82-89.

Cronbach, L.J. (1946). Response sets and test validity, *Educational and Psychological Measurement*, 6, 475-494.

Cronbach, L.J. (1950). Further evidence on response sets and test design," *Educational and Psychological Measurement*, 10, 3-31.

Ellis, B.B., & Kimmel, H.D. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology*, 77, 177-184.

Euromonitor (1994, 1997). *Consumer Europe*. London: Euromonitor International.

Fazio, R.H. (1986). How do Attitudes Guide Behavior? In: *Handbook of Motivation and Cognition. Foundations of Social Behavior*. R.M. Sorrentino & E. T. Higgins (Eds.), 204-243.

Forsman, L. (1993). Giving Extreme Responses to Items in Self-Esteem Scales: Response Set or Personality Trait? *European Journal of Psychological Assessment*, 9(1), 33-40.

Greenleaf, E.A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, 29(2), 176-188.

Greenleaf, E.A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, 56, 176-188.

Grimm, S.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.L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, 69(3), 192-203.

Hofstede, G.H. (1980). *Culture's consequences International differences in work-related values*. Beverly Hills, CA, Sage.

Hofstede, G.H. (2001). *Culture's consequences comparing values, behaviors, institutions, and organizations across nations*. Second Edition. Sage Publications. Thousand Oaks, California.

Hui, C.H., & Triandis H.C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49, 253-260.

Hui, C.H., & Triandis H.C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20(3), 296-309.

Jackson D.N., & Messick S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55, 243-252.

Johnson, T.P., & Van de Vijver, F.J.R. (2003). Social desirability in cross-cultural research. In: *Cross-cultural survey research*. J.A. Harkness, F.J.R. Van de Vijver, & P.Ph. Mohler (Eds.), 195-204.

Kahle, L.R. (1983). *Social values and social change: Adaptation to life in America*. New York: Praeger Publishers.

Knowles, E.S., & Nathan K.T. (1997). Acquiescent Responding in Self-Reports: Cognitive Style or Social Concern. *Journal of Research in Personality*, 31, 293-301.

Lee, C & Green R.T. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of International Business Studies*, 2, 289-305.

Marín, G., Gamba R.J., & Marín B.V. (1992). "Extreme response style and acquiescence among Hispanics. The role of acculturation and education. *Journal of Cross-Cultural Psychology*, 23(4), December, 498-509.

Marsh H.W., & Byrne B.M. (1993). Confirmatory factor analysis of multi-group-multi-method self-concept data: Between group and within-group variance constraints," *Multivariate Behavioral Research*, 28, 313-349.

Messick, S. (1968). Response sets. In: D.L. Sills (Ed.) *International encyclopedia of the social sciences*, 492-496. New York: Macmillan.

Messick, S. (1991). Psychology and methodology of response styles. In: *Improving inquiry in social science: A volume in honor of Lee J. Cronbach*. Snow, R.E. and D.E. Wiley, 161-200.

Mick, D.G. (1996). Are Studies of Dark Side Variables Confounded by Socially Desirable Responding? The Case of Materialism. *Journal of Consumer Research*, 23(2), 106-119.

Paulhus, D.L. (1991). Measurement and control of response bias. In: *Measures of Personality and Social Psychological Attitudes*, Volume 1. San Diego, CA: Academic Press Inc.

Peterson, R.A, G. Albaum and R.F. Beltramini (1985), "A meta-analysis of effect sizes in consumer behavior experiments," *Journal of Consumer Research*, 12(June), 97-103.

- Ray, J.J. (1983). Reviving the problem of acquiescent response bias. *The Journal of Social Psychology*, 121, 81-96.
- Schuman, H., & Presser S. (1996). *Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context*. Thousand Oaks, London: Sage Publications.
- Smith, P.B. (in press). Acquiescent Response Bias as an Aspect of Cultural Communication Style. *Journal of Cross-Cultural Psychology*.
- SouEIF, M.I. (1968). Extremeness, indifference and moderation response sets: a cross-cultural study. *Acta Psychologica*, 28, 63-75.
- Steenkamp, J.B.E.M. & Baumgartner H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, 25(June), 78-90.
- Ten Berge, J.M.F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research*, 34(1), 89-102.
- Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, 96, 506-520.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.
- Tucker, L.R. (1951). A method for synthesis of factor analysis studies. Personnel Research Section Report No. 984. Washington, DC: Department of the Army.
- Van de Vijver, F.J.R., & Leung K. (1997). Methods and data analysis for cross-cultural research. In: *Handbook of Cross-Cultural Psychology, 2nd Ed., Volume I. Theory and Method*, Berry, J.W., Y.H. Poortinga and J. Pandey (Eds.), 257-300. Boston: Allyn and Bacon.

Van Hemert, D.A., Van de Vijver, F.J.R., Poortinga, Y.H., & Georgas, J. (2002). Structure and score levels of the Eysenck Personality Questionnaire across individuals and countries. *Personality and Individual Differences*, 33, 1229-1249.

Watkins D., and S. Cheung (1995). Culture, gender, and response bias. An analysis of responses to the self-description questionnaire. *Journal of Cross-cultural Psychology*, 26(5), 490-504.

Williams, J. (1991). Constant questions or constant meanings? Assessing intercultural motivations in alcoholic drinks. ESOMAR International Research Ljubljana, February, 159-175.

Table 1 Acquiescence index

Domain	Item content	N items	Greece	Italy	Spain	France	UK	Germany	F	η^2	
Cooking											
1	Domain related	health	4	.913 ^{b,d}	.832 ^d	.490			265.64	.24	
2		like cooking	6	.191 ^{b,d}	-.031	-.024			40.43	.05	
3		calorie consumption	2	.107 ^{b,d}	-.057	-.141			17.04	.02	
4	Product specific	butter	4	.436 ^{b,d}	.267 ^d	.087			25.75	.07	
5		butter	2	.951 ^{b,d}	.816	.787			5.12	.02	
6		margarine	6	.877 ^{b,d}	.603 ^d	.465			127.78	.30	
Washing											
7	Domain related	doing laundry	3	.681 ^{b,c,d,e,f}	.542 ^{d,e,f}	.546 ^{d,e,f}	.426 ^c	.097	.341 ^c	49.07	.12
8		brand loyalty	2	.381 ^{b,d}	.162	.297	.158	.260	.314	4.82	.01
Personal care											
Electric shaving											
9	Domain specific	shaving: emotional	12		-.016 ^{d,e,f}	-.067 ^{d,e,f}	-.300 ^f	-.390 ^f	-.523	38.82	.14
10	Domain specific	shaving: sensorial	10		.415 ^{d,e,f}	.353 ^{d,e,f}	.027	.123	.117	17.97	.07
11	Domain specific	shaving: result	8		.326 ^{d,e,f}	.244 ^e	.136	.032	.137	6.57	.03
12	Domain specific	shaving: convenience	7		.683	.667	.673	.607	.654	1.03	.00
Blade shaving											
13	Domain specific	shaving: emotional	12		.120 ^{c,d,e,f}	-.056 ^{e,f}	-.128 ^{e,f}	-.281	-.318	58.24	.10
14	Domain specific	shaving: sensorial	10		.598 ^{d,e,f}	.562 ^{d,e,f}	.383	.391	.476	23.39	.04
15	Domain specific	shaving: result	8		.582 ^{c,d,e}	.482 ^e	.482 ^c	.335	.507 ^e	18.44	.03
16	Domain specific	shaving: convenience	7		.490 ^e	.539 ^{d,e}	.442 ^c	.334	.468 ^c	13.89	.03
17	Domain specific	shaving: irritation	4		.257 ^{c,d,e,f}	.089 ^e	.013	-.100	.037 ^c	24.63	.05
18	General values		7		.659 ^{e,f}	.675 ^{d,e,f}	.612 ^f	.597 ^f	.528	15.75	.02

Tukey HSD test $p < .05$; superscripts indicate a higher score;

a: significantly higher than Greece; b: significantly higher than Italy; c: significantly higher than Spain; d: significantly higher than France; e: significantly higher than the UK;

f: significantly higher than Germany

Table 2 Extreme response index

Domain	Item content	Number of items	Greece	Italy	Spain	France	UK	Germany	F	η^2	
Cooking											
1	Domain specific	health	4	.827 ^{b,d}	.700 ^d		.400		324.02	.28	
2		like cooking	6	.900 ^{b,d}	.667 ^d		.575		111.58	.12	
3		calorie consumption	2	.222 ^{b,d}	.190 ^d		.153		17.63	.02	
4	Product specific	butter	4	.559 ^{b,d}	.269		.278		20.24	.06	
5		butter	2	.853 ^{b,d}	.454		.409		21.96	.06	
6		margarine	6	.677 ^{b,d}	.321		.313		71.33	.19	
Washing											
7	Domain specific	doing laundry	3	.606 ^{b,c,d,e,f}	.360 ^f	.476 ^{b,f}	.421 ^f	.440 ^f	.281	30.83	.08
8		brand loyalty	2	.529 ^{b,c,d,e,f}	.274	.424 ^{b,f}	.399 ^{b,f}	.396 ^{b,f}	.256	21.60	.06
Personal care											
Electric shaving											
9	Domain specific	shaving: emotional	12		.448	.442	.384	.409	.475 ^d	2.90	.01
10	Domain specific	shaving: sensorial	10		.420 ^{d,e,f}	.374 ^{d,e,f}	.281	.266	.280	9.74	.04
11	Domain specific	shaving: result	8		.400 ^{d,e,f}	.333 ^{d,f}	.236	.249	.237	11.89	.05
12	Domain specific	shaving: convenience	7		.573 ^{d,e,f}	.571 ^{d,e,f}	.421	.425	.410	11.48	.05
Blade shaving											
13	Domain specific	shaving: emotional	12		.422 ^d	.423 ^d	.341	.374	.398 ^d	6.83	.01
14	Domain specific	shaving: sensorial	10		.476 ^{c,d,e,f}	.389 ^{d,e,f}	.274	.278	.304	39.85	.07
15	Domain specific	shaving: result	8		.471 ^{c,d,e,f}	.409 ^{d,e,f}	.349	.328	.339	19.57	.04
16	Domain specific	shaving: convenience	7		.462 ^{d,e,f}	.471 ^{d,e,f}	.308	.308	.307	33.83	.06
17	Domain specific	shaving: irritation	4		.354 ^{d,e,f}	.334 ^f	.278	.279	.241	9.00	.02
18	General values		7		.544 ^{d,e,f}	.520 ^{d,e,f}	.439 ^f	.452 ^f	.340	43.55	.06

Tukey HSD test $p < .05$; superscripts indicate a higher score;

a: significantly higher than Greece; b: significantly higher than Italy; c: significantly higher than Spain; d: significantly higher than France; e: significantly higher than the UK; f: significantly higher than Germany

Table 3 Consumer statistics and reported actual behavior

	Greece	Italy	Spain	France	UK	Germany
Number of times per week pastry [°]	1.7	1.9		1.9		
Variety (different kinds of dishes) [°]	4.4	4.4		4.9		
Textile washing products (national per capita value in Euro in 1993) ^{°°}	20.4	24.6	24.9	22.3	19.8	19.3
Number of times shaving per week if electric shaving ^{°°°}		5.2	5.5	6.0	6.1	5.9
Number of times shaving per week if blade shaving ^{°°°}		4.3	4.4	5.4	5.6	5.2
Sales of men's razors and blades (national per capita value in Euro in 1996) ^{°°°°}		3.4	4.6	6.4	5.5	5.7
Electric shavers (national per capita value in Euro in 1996) ^{°°°°}		.8	.7	1.7	1.7	3.2

- [°] Survey Kitchen studies (I)
- ^{°°} Euromonitor Ltd. (1994)
- ^{°°°} Survey Personal Care (III)
- ^{°°°°} Euromonitor Ltd. (1997)