A cross-national comparison of extreme response style measures

Robert A. Peterson  
*University of Texas at Austin*  
Pablo Rhi-Perez  
*University of Texas at Brownsville*  
Gerald Albaum  
*University of New Mexico*

Five measures of extreme response style were compared across 6,146 study participants from 36 countries: the traditional measure, a modified traditional measure, the individual standard deviation, an index of dispersion and an index of entropy. The traditional measure of extreme response style, whereby the two extreme categories of an item or rating scale are assigned a value of '1', all interior categories are assigned a value of '0' and the sum of the '1' values reflects the extent of extreme responding behaviour, performed slightly better than the other extreme response style measures examined with respect to reliability and ability to discriminate. The traditional measure of extreme response style was positively related to the variance of an attitudinal variable but unrelated to its mean. It was also related to Hofstede's cultural orientation variables of individualism-collectivism and power distance. Future cross-cultural and cross-national empirical research should systematically incorporate measures of extreme responding so that more is learned about the phenomenon and its possible effects.

**Introduction**

Response styles, which have traditionally been defined as systematic tendencies to respond to rating scales independently of the content of the scales (e.g. Paulhus 1991), have been extensively studied for more than half a century. Initially considered to be a form of measurement error to be avoided or addressed statistically, response styles were for many years investigated as manifestations of individual differences, especially inherent personality characteristics (e.g. Hamilton 1968). More recently, they have
been investigated in the context of cultural differences (e.g. Marshall et al. 1995; Byrne & Watkins 2003; Johnson et al. 2005; Harzing 2006; Roster et al. 2006; Wang et al. 2008; de Jong et al. 2008; Reardon & Miller 2012), national differences (Harzing et al. 2012), differing modes of data collection (Weijters et al. 2008) and rating scale format (Albaum et al. 2007b; Weijters et al. 2010). As such, by implication, response styles are frequently viewed as learned behaviours or the result of socialisation (e.g. Byrne & Watkins 2003).

Several distinct types of response style have been subjected to theoretical and empirical scrutiny. Depending on the researcher and discipline, these types have been termed ‘yea-saying’, ‘nay-saying’, ‘acquiescence’, ‘negativity’, ‘extreme responding’ and so forth. Of the various response style types, extreme responding – the systematic tendency of study participants to use the extreme (anchor) categories of an item or a rating scale (‘best or worst’, ‘love or hate’, ‘agree or disagree’, etc.) – has been relatively less studied and is one of the less well-understood response styles. Indeed, there seems to be some misunderstanding of extreme response style with respect to its measurement and implications. Consequently, the focus of this paper is on examining common extreme response style (ERS) measures and empirically comparing them in the context of a cross-national investigation involving 36 countries.

**Extreme response style**

**Common ERS measures**

Although there is virtual consensus regarding the concept and definition of extreme response style (ERS), it has been operationalised through several different measures. Four of these measures are briefly discussed here. A fifth will be introduced in the methodology section. As previously mentioned, ERS is the systematic tendency of study participants to use the extreme (end) categories of an item or a rating scale when responding.

The ‘traditional’ ERS operationalisation has been to code the extreme or anchor categories of an item or a rating scale as ‘1’ and the interior, non-extreme categories as ‘0’. Thus, whereas an item or rating scale might initially have the category structure ‘Agree 1 2 3 4 5 6 Disagree’, the categories would be recoded as ‘Agree 1 0 0 0 0 1 Disagree’ when measuring extreme response style. An individual study participant’s ERS score or value would then be the sum of recoded category responses, with a larger score reflecting a more extreme response style.
The traditional ERS measure is the most widely used metric in investigations of extreme response style. It is also the most direct measure of extreme response style. Examples of its calculation and application can be found in Baumgartner and Steenkamp (2001), Greenleaf (1992b), Grimm and Church (1999), Johnson et al. (2005), Marin et al. (1992) and Zax and Takahashi (1967).

Three other ERS measures that have been used are the standard deviation of a study participant’s responses to a set of items or rating scales, and two measures that reflect the extent to which a study participant uses all categories in an item or rating scale, an index of response dispersion and an index of response entropy. For example, Baumgartner and Steenkamp (2001), Chun et al. (1974), Diamantopoulos et al. (2006), Fischer (2004), Greenleaf (1992a), Kashima et al. (1992) and Watkins and Cheung (1995) have examined the individual standard deviation as a measure of extreme response style. Diamantopoulos et al. (2006) and Wyer (1969) have used an index of response dispersion, which measures the dispersion of a distribution, to indicate extreme response style, whereas Innes (1977) has used an index of response entropy, which measures redundancy in data, as a measure of extreme response style. However, it is important to note that these latter three measures are at best indirect measures of extreme response style.

The standard deviation of an individual study participant’s responses is at best an imperfect measure of extreme response style (e.g. Greenleaf 1992a) in that it reflects only the narrowness or broadness of that individual study participant’s response pattern. Moreover, it is an ‘asymmetric’ reflection of a response pattern. This means that, although relatively large values of the standard deviation may well reflect extreme responding, relatively small values reflect only the existence of a narrow range of responses. To illustrate, all of a study participant’s item or rating scale responses could be ‘1s’ and ‘2s’ or ‘5s’ and ‘6s’, which by themselves would suggest relatively extreme responding, but the standard deviation would be relatively small since the response range was narrow.

Likewise, the indices of response dispersion and entropy are at best imperfect measures of extreme response style (see the research methodology section for the formulas used to compute these indices). Values of both indices are a function of the number of times each item or rating scale category is used by a study participant, not the extent to which extreme response categories are used. As such, they are perhaps most appropriately considered indicators of the tendency to utilise all item or rating scale categories (e.g. Hui & Triandis 1989), not indicators of extreme response style.
Finally, a study by de Jong et al. (2008) in 26 countries presents a different approach to measuring ERS. This approach uses item response theory. These researchers asserted that this approach improved upon other methods by allowing different items to be differentially weighted for measuring ERS, and relaxed the assumption that the items in an ERS measure be uncorrelated.

**Culture- and gender-related ERS relationships**

Although some studies have concluded that there is no relationship between extreme response style and culture (e.g. Diamantopoulos et al. 2006), research generally appears to suggest that ‘extreme response bias is a cultural artifact’ (Moors 2003, p. 299) or a ‘cross-cultural circumstance’ (Clarke 2000, p. 146). Most conclusions regarding a relationship have been based on direct comparisons of study groups drawn from different cultures or nationalities. For example, direct comparisons of Hispanics and non-Hispanics have found that Hispanics were more likely to exhibit an extreme response style than were non-Hispanics (e.g. Hui & Triandis 1989; Marin et al. 1992; Clarke 2000). Direct comparisons of Korean and American college students (Chun et al. 1974), and Japanese and American college students (Zax & Takahasi 1967), revealed that, while Korean college students were more likely to exhibit an extreme response style than were American college students, American college students were more likely to exhibit an extreme response style than were Japanese college students. Chen, Lee and Stevenson (1995) found that a United States student group was more likely to use extreme scale values than were groups of Canadian, Japanese and Taiwanese students, whereas van Herk et al. (2004) found that a sample of Greek consumers was more likely to use extreme scale values than were consumer samples from France, Germany, Italy, Spain and the United Kingdom.

In addition, researchers have investigated the relationship between extreme response style and national-level cultural orientations such as those derived by Hofstede (2001). In particular, Johnson et al. (2005) used hierarchical linear modelling to examine possible relationships between extreme response style and four of Hofstede’s ‘cultural orientations’ – power distance, uncertainty avoidance, individualism–collectivism and masculinity–femininity – across 19 countries. They found that power distance and masculinity were positively related to extreme response style. The present study examined whether different relationships existed between the extreme response measures investigated and cultural orientation in a
context different from that of Johnson et al. As such, the Hofstede schema was used to investigate whether extreme response style in the countries studied had a cultural foundation.

Research results regarding the relationship between extreme response style and gender have been somewhat equivocal. Marin et al. (1992) found no consistent gender differences in their study of Hispanics’ versus non-Hispanics’ extreme responding behaviour. Similarly, Greenleaf (1992a, 1992b) in a study of US consumers, Clarke (2000) in a study of Australian, French, Mexican and United States college students, and Bachman and O’Malley (1984) in a study of response styles of blacks and whites, did not observe gender differences in their investigations of extreme responding. However, in a review article, Hamilton (1968) reported more extreme responding among females than males, and Johnson et al. (2005) reported a significant relationship between gender and extreme responding in their study.

If in fact significant differences in extreme response style do exist across cultures, nationalities, or, for that matter, demographic characteristics, the question becomes, ‘To what extent do differences in extreme response style impact substantive variables that are being investigated?’ More specifically, ‘What form does this impact take?’ and ‘How does extreme response style impact a substantive variable?’ These questions are addressed in the present research.

The effects of extreme responding

As an ‘empirical regularity’ (Chun et al. 1974), extreme response style is believed to contaminate rating scale data and ‘seriously distort the measurement of attitudes’ (Moors 2003, p. 277) because it results in the potential confounding of the content or substance of an item or rating scale with response behaviour. In other words, it may not be possible to distinguish between a strongly held attitude, opinion or perception and an extreme response style. This inability to distinguish between a strong substantive statement regarding some phenomenon and extreme response style begs an important issue: how many or what proportion of extreme item or rating scale responses in a set of items or rating scales are required before someone is deemed to possess an extreme response style?

Three distinct instances of distortion have been hypothesised to occur: distortion of a mean, distortion of a variance and distortion of a correlation coefficient. With respect to distortion of the mean, Cheung and Rensvold (2000, p. 189) concluded that, ‘because ERS affects numerical
scores, comparisons of means become uninterpretable’. However, while ERS may well influence the level of response – that is, the mean response – it is not clear that, contrary to common belief, ERS will in practice *systematically* bias the response level or mean response in any meaningful manner, either artificially inflating or deflating mean values. For example, even though Chen et al. (1995) obtained statistically significant differences in ERS across groups of college students from Canada, Japan, Taiwan and the US, they noted that these differences only ‘exerted a very modest influence on the degrees of difference found in comparisons of the groups’ mean ratings’ (p. 174).

The lack of substantive influence on response levels or response means that what Chen et al. (1995) observed probably resulted in part from two independent factors. First, inspection of the results of prior ERS studies revealed that, on average, somewhere between 5 and 10% of all responses to a set of items or rating scales typically fall into the extreme categories. This range is substantially less than might be expected ‘by chance’. For a five-category scale, the ‘expected’ percentage of extreme responses is 40%; for a six-category scale the expected percentage is 33.3%. Thus, the relatively small percentage of extreme responses observed in prior studies suggests that any impact may be at best modest.

Second, by definition, extreme response style consists of using extreme item or rating scale response categories. Unless a study participant systematically uses only one extreme category of an item or rating scale to represent, say, a positive or negative evaluation, in which case the response style would be something other than ERS, there should be some sort of balancing or trade-off of extreme responses, both within an individual and among individuals. Depending on the ‘mix’ of extreme categories utilised, such balancing or trade-offs should intuitively only minimally influence the response level or mean of an item or rating scale. For example, a symmetric distribution with 10% of the responses falling in the two extreme categories of an item or rating scale would not affect the overall level of response or mean because extreme responses would effectively cancel each other out.

Simultaneously, though, systematic use of extreme item or rating scale categories by study participants would be expected to add nuisance or extraneous variation to a data set and consequently artificially increase the variance of an item or rating scale. More extreme responding should be reflected in larger item or rating scale (variable) variances; less extreme responding should be reflected in smaller item or rating scale (variable) variances. Therefore, systematic differences in extreme responding across, say, cultures or countries being compared could adversely affect statistical
tests based on means, such as standard t-tests or F-tests, by either differentially inflating or deflating standard errors artificially.

Third, because correlation coefficients are based in part on variable variances, extreme responding could attenuate estimated relationships between two variables. Heide and Gronhaug (1992) conducted a series of simulations to evaluate the effect of extreme response style on the magnitude of correlation coefficients. They found that, as the incidence of extreme responding increased in a data set, the magnitudes of correlation coefficients based on the data decreased. This occurred because the increased incidence of extreme responding resulted in inflated variable variances that, in turn, degraded the correlation coefficients.

Thus, to the extent that the incidence of extreme responding is systematically and significantly different in different cultures or countries, the magnitude of within-country correlations could vary substantially due to a measurement artefact. In certain instances, though, the potential effect of extreme responding on correlation coefficients is more insidious. Because substantive responses to a set of items or rating scales may be confounded with extreme response style, reliability or validity coefficients could be artificially inflated if extreme responding behaviour, rather than item or rating scale content, consistently determined study participants’ responses.

Moreover, since the correlation coefficient is the basis of several multivariate statistical techniques (e.g. regression analysis, canonical correlation analysis, factor analysis), applications of these techniques could also be affected by extreme response style. Indeed, Arce-Ferrer and Ketterer (2003) have demonstrated the influence of extreme responding on the recovery of the factor structure underlying the self-efficacy construct, and Cheung and Rensvold (2000) have shown that extreme responding affects factor loadings and intercept terms in confirmatory factor analysis.

In brief, in the present study the potential impact of extreme responding on the mean of a multi-item attitude measure was expected to be minimal. At the same time, though, a positive relationship between the variance of the attitude measure and extreme responding was expected. Because extreme responding could artificially inflate or deflate correlation coefficients, there was no expectation as to its effect on the correlation between an attitude measure and gender.

**Research questions**

The present research sought to answer four broad questions. In the context of the particular ERS measures examined, the attitudinal and opinion
variables investigated, the study participants employed, and the countries serving as the source of the study participants:

1. To what extent are the ERS measures reliable?
2. What relationships, if any, exist among the ERS measures?
3. What are the characteristics of the ERS measures, and are there differences among the ERS measures across the countries investigated?
4. Are there relationships respectively between the ERS measures and (a) cultural orientations, (b) an attitude measure, and (c) gender?

Research methodology

Samples

The primary sample obtained consisted of 6,146 undergraduate business students (from an original sample of 6,333) from 120 universities in 36 countries. Because the US tends to be the source country for subjects in a plurality of cross-national studies, students in the US were purposely over-sampled (\(N = 2,949\)) and used as a benchmark when interpreting study results. The combined obtained aggregate sample from the other 35 countries consisted of 3,197 survey participants. A two-stage sampling design was employed in which a sample of universities was first identified; this was followed by a cluster sample of business students in each of the sampled universities. Approximately 88% of the US universities were public universities; about two-thirds of the non-US universities were public universities. Data were collected by means of self-administered questionnaires in an in-class setting, to control for possible ‘noise’ by having a common data collection environment. Only those study participants who answered all items in the questionnaire were included in the analyses, resulting in a total of 187 students being deleted.

Table 1 contains a listing of the countries from which the samples were obtained, the number of universities sampled in each country, and the number of students sampled in each country. Because prior research on extreme response style revealed that extreme responding tended to be related to demographic characteristics such as age and education, business students were purposely used as study participants in an attempt to obtain as homogeneous a sample of study participants as possible with respect to age and education (see, for example, Clarke 2000; Reynolds et al. 2002). Although there were relatively large differences in the country sample sizes, these differences are comparable to those observed
Table 1  Countries included in the study*

<table>
<thead>
<tr>
<th>Country</th>
<th>Universities</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>France</td>
<td>2</td>
<td>150</td>
</tr>
<tr>
<td>New Zealand</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Norway</td>
<td>2</td>
<td>183</td>
</tr>
<tr>
<td>Philippines</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td>Senegal</td>
<td>3</td>
<td>109</td>
</tr>
<tr>
<td>Singapore</td>
<td>2</td>
<td>117</td>
</tr>
<tr>
<td>South Korea</td>
<td>2</td>
<td>86</td>
</tr>
<tr>
<td>Spain</td>
<td>2</td>
<td>174</td>
</tr>
<tr>
<td>Thailand</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>Tunisia</td>
<td>3</td>
<td>212</td>
</tr>
<tr>
<td>Turkey</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>United States</td>
<td>58</td>
<td>2,949</td>
</tr>
<tr>
<td>Vietnam</td>
<td>2</td>
<td>54</td>
</tr>
</tbody>
</table>

*The numbers in parentheses are, respectively, the number of universities and students sampled in each country. Thus, in Argentina, one university served as the source of 59 study participants.

in other cross-national research studies (e.g. van Hemert et al. 2002). In general, whereas the sample cannot be considered globally representative, it included individuals with heterogeneous cultural backgrounds and a common educational focus.

In addition to the primary sample, 137 business students from two universities in France, one in Spain, and one in the US were used to evaluate the longitudinal stability of the ERS measures compared. This was accomplished by means of a two-week test–retest correlation.

Measurement instrument

The questionnaire used contained 27 Likert-type items, six demographic questions and three questions – academic classification, major field of study, and citizenship – used to screen potential survey participants to ensure that the sample was limited to only undergraduate business students from the respective countries. All students were citizens of the country where they were sampled. Each of the 27 Likert-type items consisted of a declarative statement and a balanced six-category ‘strongly agree’ to ‘strongly disagree’ rating scale. Scale categories were labelled numerically from 1 (‘Strongly Agree’) to 6 (‘Strongly Disagree’). Only the endpoints of the rating scale were labelled. The items were derived from several sources and pretested on a sample of undergraduate business students in the US; the pretest included a qualitative evaluation of item understandability and an assessment of administrative ease.

The questionnaire was developed in English and subsequently translated into Chinese, French, German, Spanish and Vietnamese, usually by
professors in the countries where data were collected, then back-translated for interpretability. Because many of the study participants were from English-speaking countries or educational institutions, or spoke English or one or more of the languages into which the questionnaire was translated (i.e. Filipinos speak Spanish and Tunisians speak French), there was no need to translate the questionnaire into other languages. While some students were queried in a language that was a second language to them (e.g. English), this was not deemed problematic as they were being instructed in the second language.

**Study variables**

There were six study (independent or moderating) variables and five ERS measures. The 27 Likert-type items consisted of items measuring two constructs: attitude towards capitalism and business ethicality opinions. Attitude towards capitalism (ATC) was measured using four items:

1. capitalism advocates the work ethic and the free market mechanism
2. a free society can exist under capitalism
3. efficient management can only be achieved through capitalism
4. capitalists are entitled to the reward of profits because they assume the risk of loss.

Given the different cultures represented in the sample, it was assumed that the attitudinal variable would produce broad and perhaps heterogeneous response distributions.

The four items were summed to produce a multi-item scale; this scale constituted one dependent variable. ATC scale scores could theoretically range from 4, indicating a positive attitude towards capitalism, to 24, indicating a negative attitude towards capitalism. The mean for the total (primary) sample was 12.1; the standard deviation was 3.7; the US mean and standard deviation were 11.9 and 3.8. Coefficient alpha was 0.68 for the total sample and 0.77 for the US sample. Given the nature of the investigation and the study participants (business students) and countries (communist, capitalist, socialist) sampled, ATC was deemed an appropriate dependent variable.

The other study variables were four Hofstede-generated (Hofstede 2001) variables – power distance, uncertainty avoidance, individualism–collectivism and masculinity–femininity – variables studied by Johnson et al. (2005), and gender. As stated previously, the evidence regarding
the relationships between these culture-dependent variables and extreme response style is equivocal, especially cross-nationally. Therefore, the present intent was to provide additional insights into possible relationships between culture and extreme responding.

The five ERS measures were calculated on the 23 remaining items. These items covered a variety of ethics-related opinions, as illustrated by the examples shown in Table 2. The topic of business ethics has taken on increased research interest during the past few years as a result of the legal and ethical lapses by executives of such companies as Enron, WorldCom, Arthur Andersen and Tyco. This interest has extended to research across countries. For example, Axinn et al. (2004) studied the perceived importance of ethics and social responsibility among respondents in the US, Malaysia and Ukraine. More recently, Albaum et al. (2007a) looked at the ethicality of future business leaders in 23 countries. In the present study, although the 253 \( \frac{(23 \times 22)}{2} \) inter-item correlations ranged from -0.01 to +0.62, most were under 0.1 and many were close to zero. Thus, the ERS measures were derived independently of the dependent variables with a mean absolute inter-item correlation that compared favourably with the mean absolute values that were calculated for a 16-item scale explicitly developed to measure extreme response style (e.g. Greenleaf 1992b; Clarke 2000). Hence, it would seem that the likelihood of confounding the content of ATC with a measure of extreme responding was minimised such that responses to the 23 items could serve as meaningful data for computing the five ERS measures being compared.

Table 2  Examples of Likert scale items relevant to ethics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The growth of the large multinational corporation has led to a decline in ethical behaviour in business.</td>
</tr>
<tr>
<td>2.</td>
<td>Business managers must often engage in behaviours that I consider to be unethical.</td>
</tr>
<tr>
<td>3.</td>
<td>In order to succeed in business, it is often necessary to compromise one's ethics.</td>
</tr>
<tr>
<td>4.</td>
<td>Top business executives should state in no uncertain terms that unethical behaviours in their companies will not be tolerated.</td>
</tr>
<tr>
<td>5.</td>
<td>If a manager in a company is discovered to have engaged in unethical behaviour that results primarily in personal gain (rather than corporate gain), he or she should be terminated or fired.</td>
</tr>
<tr>
<td>6.</td>
<td>If a manager in a company is discovered to have engaged in unethical behaviour that results primarily in corporate gain (rather than personal gain), he or she should be terminated or fired.</td>
</tr>
<tr>
<td>7.</td>
<td>Business behaviour that is legal is ethical.</td>
</tr>
<tr>
<td>8.</td>
<td>Within a business firm, the ends justify the means.</td>
</tr>
</tbody>
</table>
ERS measures

Of the five ERS measures, the first one calculated was the traditional, direct measure of extreme responding ("EXR"). Each of the two end points or anchor categories (i.e. '1' and '6') for the 23 items was coded '1', whereas the remaining inner categories ('2', '3', '4' and '5') were coded '0'. The recoded responses were summed to produce EXR values that could theoretically range from 0 to 23, with larger values representing more extreme responding.

The second ERS measure calculated was a modified version of EXR ('MOD'). Rather than coding only item categories '0' or '1' to represent the existence of extreme responding, MOD was calculated by assigning every item category a non-zero numeric value. Specifically, whereas the item categories were initially coded 'Agree 1 2 3 4 5 6 Disagree', MOD was derived by recoding the categories such that a revised item category structure 'Agree 3 2 1 1 2 3 Disagree' was created. MOD was designed to serve as a holistic measure that incorporated both extreme responding as well as mid- or inner-category responding; its values could theoretically range from 23 to 69.

The third ERS measure calculated was the standard deviation (ISTD) of each individual study participant's responses across the 23 items. ISTD reflects only the narrowness or broadness of an individual study participant's response pattern. Moreover, it is an 'asymmetric' reflection of a study participant's response pattern. Relatively large values of ISTD may well reflect extreme responding, but relatively small values only reflect a narrow range of responses. ISTD values could range from 0.0 to 2.56 in the present situation.

The fourth ERS measure, an index of response dispersion (DIS), was calculated (Wyer 1969) as DIS = \( k(m^2 - \sum_{i=1}^{k} c_i^2)/m(k-1) \), where \( k \) is the number of item categories (6 here), \( m \) is the total number of items (23 here), and \( c_i \) is the number of times item category \( i \) was used by an individual study participant. Thus, the more item categories used, the larger the value of DIS. However, analogous to ISTD, DIS does not necessarily reflect extreme responding; by definition and intent, it merely measures response dispersion. Its values could theoretically range from 0.0 to 1.0.

The fifth ERS measure produces an estimate of entropy, response randomness or information (ENT). It was calculated (Shannon & Weaver 1963; Peterson & Sharma 1977) as \( \text{ENT} = -H \sum_{i=1}^{k} p_i \log p_i \), where \( H \) is a scaling constant (ignored here), \( p_i \) is the proportion of times item category \( i \) was used by an individual study participant, and \( k \) is the number of categories in an item or rating scale. The greater the value of ENT, the
more random (less predictable) the response pattern of an individual study participant; the maximum value of ENT would occur with a uniform response pattern (all item or rating scale categories used equally). In this sense it is analogous to the index of dispersion. In the present situation, values of ENT could range from 0.0 to 1.76.

As previously mentioned, ISDT, DIS and ENT are not technically measures of ERS. Even so, they are included because they have been considered to be such measures and in part they provide a comparative context.

Results and discussion

ERS measure reliability and relationships

The first two research questions addressed the reliability of the ERS measures examined and whether any relationships existed among them. Table 3 presents reliability coefficients (longitudinal stability or test–retest correlations) and intercorrelations among the ERS measures. The diagonal of Table 3 contains the reliability coefficients. Based on a multinational sample of 137 study participants, the two-week longitudinal reliabilities of the ERS measures appear reasonable, although the reliability coefficients for the indexes of dispersion and entropy are lower than those for the other three measures. The traditional ERS measure had the largest reliability coefficient (0.74), slightly larger than those for the modified measure (0.72) and the individual standard deviation (0.71). With the exception of the reliability coefficients for the indexes of dispersion and entropy, the reliabilities were generally similar to a one-week test–retest correlation of 0.74 reported by Merrens (1970) and a 15-day test–retest correlation (0.78) reported by Berg (1953; Berg & Collier 1953) in studies of extreme

Table 3  ERS measure correlations

<table>
<thead>
<tr>
<th>ERS measure</th>
<th>EXR*</th>
<th>MOD</th>
<th>ISTD</th>
<th>DIS</th>
<th>ENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXR**</td>
<td>(0.74)***</td>
<td>0.92</td>
<td>0.81</td>
<td>0.24</td>
<td>0.34</td>
</tr>
<tr>
<td>MOD</td>
<td>0.93</td>
<td>(0.72)</td>
<td>0.85</td>
<td>0.33</td>
<td>0.41</td>
</tr>
<tr>
<td>ISTD</td>
<td>0.82</td>
<td>0.85</td>
<td>(0.71)</td>
<td>0.60</td>
<td>0.68</td>
</tr>
<tr>
<td>DIS</td>
<td>0.13</td>
<td>0.22</td>
<td>0.51</td>
<td>(0.59)</td>
<td>0.96</td>
</tr>
<tr>
<td>ENT</td>
<td>0.24</td>
<td>0.32</td>
<td>0.60</td>
<td>0.96</td>
<td>(0.61)</td>
</tr>
</tbody>
</table>

* Entries above the diagonal are calculated on US data (n = 2,949)
** Entries below the diagonal are calculated on all data (n = 6,146)
*** Entries on diagonal are test–retest reliabilities (n = 137)
response style. Thus extreme responding, as investigated here, appears to be relatively consistent over time.

Correlation coefficients below the diagonal in Table 3 were computed on all 6,146 study participants. Correlation coefficients above the diagonal in Table 3 were computed on 2,949 American study participants for comparison purposes. In general, the corresponding correlations in the two matrices were fairly consistent. The only noticeable differences were those involving the index of dispersion; for three out of four correlations involving the index of dispersion, the US correlations were larger than those for the total sample (which includes the US).

The traditional measure of extreme response style (EXR) and the modified measure (MOD) were highly correlated in the total sample (0.93) as well as the US sample (0.92), as were the index of dispersion (DIS) and the index of entropy (ENT) (0.96/0.96). ISTD was related to all four measures, but to a lesser degree than the respective pairs of measures. To explore the measure relationships in more detail, factor analyses were conducted for the total sample and the US sample using the five ERS measures as input variables. As might be expected from the patterns of the correlation coefficients, a varimax factor solution identified two distinct and virtually identical factors for the two samples, as shown in Table 4. The first factor includes the EXR and MOD measures, whereas the second factor includes the DIS and ENT measures. Thus, as expected, EXR-MOD and DIS-ENT appeared to be measuring different things, and ISTD appeared to have common properties with both pairs of measures.

ERS measure characteristics

The third research question inquired about the characteristics of the ERS measures and whether differences in the measures existed across countries. Table 5 presents the minimum mean, the maximum mean and the average mean value for each of the five ERS measures at the country level, as well as maximum mean/minimum mean ratios and F-ratios based on one-way analyses of variance for the five measures at the individual level. Additionally, ERS means of US study participants are presented for reference purposes. In general, based on maximum mean/minimum

<table>
<thead>
<tr>
<th>Sample</th>
<th>Loading of ISTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>All respondents</td>
<td></td>
</tr>
<tr>
<td>Factor 1 (EXR-MOD)</td>
<td>0.85</td>
</tr>
<tr>
<td>Factor 2 (DIS-ENT)</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Table 4 Factor loadings of ISTD measure on ERS factors
Table 5  Country-level ERS measure characteristics

<table>
<thead>
<tr>
<th>ERS measure</th>
<th>US mean</th>
<th>Maximum mean</th>
<th>Minimum mean</th>
<th>Average mean</th>
<th>Maximum/minimum</th>
<th>F-ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXR</td>
<td>5.02</td>
<td>11.46 (Morocco)</td>
<td>1.74 (Ireland)</td>
<td>5.24</td>
<td>3.01</td>
<td>28.9</td>
</tr>
<tr>
<td>MOD</td>
<td>41.16</td>
<td>51.92 (Morocco)</td>
<td>34.00 (Ireland)</td>
<td>41.69</td>
<td>1.23</td>
<td>27.7</td>
</tr>
<tr>
<td>ISTD</td>
<td>1.36</td>
<td>1.69 (Senegal)</td>
<td>1.03 (Bolivia)</td>
<td>1.37</td>
<td>1.64</td>
<td>19.4</td>
</tr>
<tr>
<td>DIS</td>
<td>0.88</td>
<td>0.93 (China)</td>
<td>0.77 (Bolivia)</td>
<td>0.90</td>
<td>1.21</td>
<td>10.3</td>
</tr>
<tr>
<td>ENT</td>
<td>2.11</td>
<td>2.27 (Turkey)</td>
<td>1.73 (Greece)</td>
<td>2.10</td>
<td>1.21</td>
<td>10.6</td>
</tr>
</tbody>
</table>

*F-ratio based on one-way ANOVA with individual-level data

mean ratios and F-ratios, the traditional measure of extreme response style appears to be the most discriminating measure across the countries studied.

Conversion of the specific means of the traditional ERS measure in Table 5 into percentages reveals that, on average, the overall sample percentage incidence of extreme category responding, 23%, was significantly less than the expected ('chance') percentage value of 33%. Indeed, study participants in only five countries – Morocco, Senegal, Honduras, Columbia and Tunisia – exhibited extreme responding at or greater than the expected 33% level. The extreme responding percentage for US study participants, 22%, was not significantly different from that for non-US study participants, 24%. Study participants from Morocco exhibited the largest extreme responding percentage, 50%, whereas study participants from Ireland exhibited the smallest extreme responding percentage, 8%.

In addition to the values in Table 5, the skewness and kurtosis of the ERS measure distributions were computed for each country. Using a threshold of ±1.5 as an indication of a substantial departure from normality, DIS values were negatively skewed and platykurtic, whereas ENT values were platykurtic regardless of country. The distributions of ISTD and MOD were relatively symmetric across countries. The distributions of EXR values were somewhat skewed and platykurtic in eight countries (Belgium, Bolivia, Hong Kong, Netherlands, Norway, Singapore, Spain and the UK). Although no systematic skewness and kurtosis EXR patterns were detected in these eight countries, the fact that significant distributional differences did occur in more than one-fifth of the countries studied suggests the need for further investigation. Importantly, the fact that none of the extreme response measure values was normally distributed needs to be taken into account when interpreting individual as well as country values.
ERS measures, culture, capitalism attitude and gender

The fourth research question inquired about possible relationships between the ERS measures and cultural orientation, attitude (e.g. attitude towards capitalism) and gender. Table 6 presents country-level correlations between the means of each of the ERS measures and four of Hofstede’s cultural orientation variables: power distance, uncertainty avoidance, individualism–collectivism and masculinity–femininity.

Only two of the 20 correlations were statistically significant. EXR was positively associated with power distance and negatively associated with individualism–collectivism. Although the positive correlation observed between EXR and power distance corroborated a finding of Johnson et al. (2005), there were no systematic relationships between the ERS measures and Hofstede’s cultural variables. Consequently, whatever differences in ERS measures existed were not necessarily culture based.

Table 7 presents correlations between the ERS measures and the mean and variance of the attitude towards capitalism (ATC) variable. As expected, there was no meaningful relationship between the ERS measures and the ATC mean at the country level. Although the ENT-ATC mean correlation was statistically significant, its magnitude reflected only 16% shared or explained variance. However, also as expected, there were significant pairwise relationships between three of the ERS measures – EXR, MOD and IST – and the variance of the attitude towards capitalism variable. As reflected by these relationships, the greater the incidence of extreme responding in a country, the larger the variance in the attitude of study participants in that country towards capitalism. Thus, extreme responding appears to influence the variance but not the mean of the attitude towards capitalism variable.

Table 6  Country-level correlations between ERS measures and Hofstede variables*

<table>
<thead>
<tr>
<th>ERS measure</th>
<th>Hofstede variable</th>
<th>Power distance</th>
<th>Uncertainty avoidance</th>
<th>Individualism–collectivism</th>
<th>Masculinity–femininity</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXR</td>
<td></td>
<td>0.35**</td>
<td>0.22</td>
<td>-0.35**</td>
<td>0.01</td>
</tr>
<tr>
<td>MOD</td>
<td></td>
<td>0.31</td>
<td>0.18</td>
<td>-0.33</td>
<td>0.02</td>
</tr>
<tr>
<td>ISTD</td>
<td></td>
<td>0.20</td>
<td>0.11</td>
<td>-0.23</td>
<td>-0.06</td>
</tr>
<tr>
<td>DIS</td>
<td></td>
<td>-0.15</td>
<td>-0.04</td>
<td>0.12</td>
<td>-0.13</td>
</tr>
<tr>
<td>ENT</td>
<td></td>
<td>-0.09</td>
<td>-0.01</td>
<td>0.08</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

*n = 36; **p < 0.05
Table 7  Country-level correlations between ERS measures and ATC mean and variance*

<table>
<thead>
<tr>
<th>ERS measure</th>
<th>ATC mean</th>
<th>ATC variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXR</td>
<td>-0.06</td>
<td>0.56**</td>
</tr>
<tr>
<td>MOD</td>
<td>-0.07</td>
<td>0.54**</td>
</tr>
<tr>
<td>ISTD</td>
<td>-0.10</td>
<td>0.49**</td>
</tr>
<tr>
<td>DIS</td>
<td>-0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>ENT</td>
<td>-0.40**</td>
<td>0.13</td>
</tr>
</tbody>
</table>

* n = 36; ** p < 0.05

Although three of the five ERS measures correlated significantly with gender for the total sample, EXR (0.05), MOD (0.06) and ISTD (0.05), significance was due more to the relatively large sample size (cf. Mohr 1990, p. 74) than to the magnitudes of the correlations. Thus, while there was a tendency for males to respond more extremely than females, there were no substantive relationships between the ERS measures and gender. This finding tends to corroborate previous findings regarding the absence of a relationship between extreme responding and gender (e.g. Bachman & O’Malley 1984; Greenleaf 1992a, 1992b; Marin et al. 1992; Clarke 2000).

The relationship between ATC and gender

The (zero order) correlation between attitude towards capitalism and gender was -0.17 for the sample of 6,146 study participants; male study participants were more favourably disposed towards capitalism than were female study participants. To investigate the possible effect of extreme responding on this correlation, partial correlations were computed between attitude towards capitalism and gender controlling for extreme response style. Regardless of which ERS measure was used as the control, there was no effect of extreme responding on the correlation. This was also true for the sample of US study participants; the zero order and partial correlations were identical at -0.21. The absence of a partialling effect was not unexpected given the respective correlations between ERS and gender and ERS and ATC.

There were, however, differences among the magnitudes of the partial correlations within countries. For example, partial correlations ranged from 0.01 for Greece to -0.40 for Denmark when the traditional ERS measure (EXR) was used as a control variable. This suggests the need to examine the possible effect of extreme responding within each of the countries included in a study.
Conclusions

Analogous to other empirical studies of extreme response style, conclusions drawn from this one must be tempered by the nature of the subject matter, the particular study participants employed, the countries sampled, and the specific operationalisations of the ERS measures and major variables. Moreover, although the ERS measures were computed on items that possessed low inter-correlations and were independent of the other variables employed in the study, there still remains the possibility of response style-item content confounding for the 23 ethics-related items.

Even so, the results of the present study merit close scrutiny regarding their implications for cross-cultural and cross-national research. In brief, the present results support prior conclusions that extreme response style is an empirical regularity that is both reliable, thus providing a ‘yes’ answer to Research Question 1, which was ‘To what extent are the ERS measures reliable?’ and possesses the potential to differentially influence research findings and conclusions (e.g. Chun et al. 1974). Results also revealed that EXR and MOD were highly correlated as were DIS and END, thus providing an answer to Research Question 2, which asked, ‘What relationships, if any, exist among the ERS measures?’

Research Question 3 was ‘What are the characteristics of the ERS measures, and are there differences among the ERS measures across the countries investigated?’ Of the five ERS measures that were compared in the present study, the traditional measure, wherein extreme categories of an item or rating scale are coded as ‘1’ and the remaining inner categories are coded as ‘0’ and the recoded responses summed to produce a score for each study participant, was found to be marginally more stable and discriminating than the other four ERS measures across the countries investigated. However, despite differences in the means, variances and distributions of the traditional ERS measure across countries, conversion of the means obtained in the present study into percentages revealed that the level of extreme responding generally did not exceed what might be expected by chance.

Research Question 4 asked, ‘Are there relationships respectively between the ERS measures and (a) cultural orientation, (b) an attitude measure, and (c) gender?’ While extreme response style appeared to affect the variance of the attitude towards capitalism variable, it did not significantly affect the mean of this variable or its relationship with gender. Moreover, although three ERS measures (EXR, MOD, ISTD) correlated with gender, the correlation coefficients were less than 0.10, and significance was most probably due to the large sample size.
Although published research on extreme responding encompasses more than six decades, much is still not known about this phenomenon. For example, the possible effect of language on extreme responding pointed out by Stening and Everett (1984) needs to be explored further, especially since Harzing (2006) has reported that English-language questionnaires elicit a high level of middle responses (i.e. central tendency), whereas questionnaires in a non-English respondent’s (native) language result in more extreme responses. Consequently, more methodological and programmatic research is required regarding the effects of extreme response style on the outcomes of cross-cultural and cross-national investigations.

Because such effects may well be complex, and systematically and differentially nested within cultures or countries, broad generalisations about extreme responding are not warranted at this time. Until non-student cross-national or cross-cultural samples have been studied with a variety of attitudinal, perceptual and other individual difference variables, the practical effects of extreme response style must be regarded as tentative. Finally, it is necessary to study the difference between extreme response style and honest and accurate answering with extreme responses. As a first step, researchers should incorporate extreme response-style metrics in all survey-based cross-national and cross-cultural studies so that actuarial-type data can be obtained on a variety of heterogeneous variables and individuals.

References


**About the authors**

Robert A. Peterson is Associate Vice President for Research and Director of the IC2 Institute at The University of Texas at Austin and holds the Stuart Chair in the McCombs School of Business. He received his PhD
from the University of Minnesota in 1970. The author or co-author of more than 170 refereed articles and books, he is a former editor of the Journal of Marketing Research, the Journal of the Academy of Marketing Science and the current co-editor of the AMS Review.

Pablo Rhi-Perez is Associate professor of Marketing and Entrepreneurship in the School of Business, the University of Texas at Brownsville. His research interests are in the field of Mexican American cross-cultural consumer behaviour and minority entrepreneurship. Dr Rhi-Perez is currently Director of the commercialization Center at the University of Texas at Brownsville. A bi-national economic development zone and substitution of imported components and parts by the regional maquiladora industry are some of the working projects under the Center.

Gerald Albaum is Research Professor at the Robert O. Anderson Schools of Management at the University of New Mexico. He also is a senior Research Fellow at the IC2 Institute, University of Texas at Austin, and is Professor Emeritus of Marketing at the University of Oregon. He received his PhD in 1962 from the University of Wisconsin-Madison. He is the author or co-author of numerous books and articles in refereed journals and conference proceedings. His writings deal with issues in research methods, international marketing activities and direct selling.

Address correspondence to: Gerald Albaum, Anderson Schools of Management, MSC05 3090, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA.

Email: albaum@unm.edu
Copyright of International Journal of Market Research is the property of Warc LTD and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.