

Extreme Response Style: A Meta-Analysis

John H. Batchelor
University of West Florida

Chao Miao
Wilkes University

Extreme response style (ERS) refers to the tendency to prefer responding using extreme endpoints on rating scales. We use meta-analysis to summarize the correlates of ERS. Our findings present how one's tendency to engage in extreme responding is related to demographic variables (e.g., age, gender, education, and race), intelligence, acquiescence, and number of points and items in a scale. We also identified a non-linear relationship between age and extreme responding. Thus, this article should be read by anyone using Likert type scales when using data from a diverse set of individuals.

INTRODUCTION

Extreme response style (ERS) refers to the tendency to prefer responding using extreme endpoints on rating scales; this tendency is independent of item content (Greenleaf, 1992). Thus, one high on an ERS would tend to select either high or low values on a Likert scale, but one low on ERS would tend to select more mid-scale values. Thus, ERS, sometimes called extreme responding (Lau, 2007), is the opposite of the central tendency response style tendency (Naemi, Beal, & Payne, 2009). ERS was also viewed as a content irrelevant factor that can influence an individual's response to rating scales (Cronbach, 1946, 1950). A response style, such as ERS, is typically viewed as stable across time and situations (Jackson & Messick, 1958; Lau, 2007).

Response Style Versus Response Set

ERS and any other response style can be contrasted with a response set. Response set refers to “systematic ways of answering which are not directly related to the question content, but which represent typical behavioral characteristics of the respondent” (Oskamp, 1977, p. 37). Often, a response set involves dissimulation such as efforts by the respondent to respond in a socially desirable manner (Rorer, 1965). Additionally, response set can represent “temporary response tendencies” (Paulhus, 1991) such as temporal mood variations.

Response styles may be viewed as “statistical nuisances” and should be treated as such, through adequate controls (Couch & Keniston, 1960). The presence of ERS may result in bias with respect to construct measurement and associations with other variables (Moors, 2004). ERS can be particularly damaging to measurement when scales lack item balance with respect to item direction. For example, a conscientiousness scale in which all items are phrased such that high ratings always indicate high conscientiousness would be unbalanced. Consider two respondents with equal levels of conscientiousness

(perhaps slightly above the mean; a 5 on a 7 point Likert scale) but differing levels of ERS. On the 7-point Likert scale, the respondent who tends to have few ERS tendencies will score an average of 5 on each item. However, the respondent prone to extreme responses scores will score an average of closer to 7 on each item. Even when items in scales are balanced, ERS adds constructs-irrelevant variance to the ratings scales. Also, some measures cannot be readily balanced. For example, in job analysis, it would be cumbersome to balance ratings scales concerning importance of skills or frequency of task performance.

Because ERS increases construct-irrelevant variance, it inflates within group variance and this reduces statistical power. Also, the construct-irrelevant ERS variance will reduce the magnitudes of relationships among variables. To understand the magnitude of this decrease, Lau (2007) found that ERS reduced the explained variance in his study from 69.5% to 53.5%.

McDaniel, Psotka, Legree, Yost, and Weekley (2011) demonstrated that ERS can be particularly damaging in measures that are consensually scored. Consensual scoring is used when response options cannot be easily declared to be correct or incorrect. This is typically the case in situational judgment test. In consensual scoring, expert judges are often used to identify best response and on Likert-format response scale, scores are expressed as deviations of the respondent's Likert ratings from the expert judge ratings. The expert judge ratings are more likely to be found near the center of the Likert scale than at the extremes of the scales. This results in respondents with ERS tendencies to score lower, on average, than respondents with limited or no ERS tendencies. In the prediction of job performance, McDaniel et al. judged the ERS variance in situational judgment tests to be criterion-irrelevant. When that variability was subject to statistical corrections, the correlations with job performance sharply increased.

Current State of Literature on ERS

Although ERS has been investigated since the early 1950's (Berg, 1953), the effects of ERS are generally ignored by researchers (Hamilton, 1968; Paulhus, 1991). In addition, recent research suggested that there is little knowledge on the correlates of ERS and in-depth analyses on this area is sorely needed (e.g., Bockenholt, 2012; Naemi, Beal, & Payne, 2009; Wetzel, Lüdtke, Zettler, & Böhnke, 2015). Despite decades of research, Wetzel et al. (2015) lamented that the correlates of ERS still remain largely unclear. Further, the empirical evidence regarding the correlates of ERS is mixed; suggesting the need for meta-analysis and explain the inconsistent findings that exist in the current ERS literature. Here we provide an overview of major themes in the ERS literature. We then offer a meta-analysis of the correlates of ERS.

Research suggests that there may be demographic differences with regards to ERS. For instance, Bachman, O'Malley, and Freedman-Doan (2010) found that Blacks and Hispanics tend to respond more extremely than Whites and Asians. Consistent with this finding, McDaniel et al. (2011) demonstrated that adjusting for ERS substantially reduced White-Black mean score differences in two situational judgment measures. The literature regarding age and ERS is mixed. Some studies found that as individuals age they engage in less ERS (Bachman et al., 2010; Greenleaf, 1992; Iwawaki et al., 1969), but others found ERS to increase as one ages (Clarke, 2000). Still others found no differences with reference to age (Johnson et al., 2005). Finally, Hamilton (1968) and Lau (2007) argued that the relationship between ERS and age may be curvilinear, thus increasing with age up to a point and then decreasing after one reaches a certain age. Although the ERS sex difference literature is mixed, the sex differences tend to be small (Bachman & O'Malley, 1984b; Zuckerman, Oppenheimer, & Gershowitz, 1965; Greenleaf, 1992).

Most arguments for why demographic groups differ in relation to ERS center on how these groups tend to view the world around them (Neuringer, 1961). For instance, Bachman and O'Malley (1984b) discussed how Blacks tend to use "unqualified terms" when expressing themselves; in contrast Whites use of "caution and inhibition." With regard to culture, ERS is argued to be less pervasive in collectivist cultures than individualistic cultures (Triandis, 1995) and more evident in masculine than feminine cultures (Johnson et al., 2005). Further, ERS is considered by some to be associated with the desire to achieve "clarity, precision, and decisiveness" (Johnson et al., 2005). For those engaging in cross-cultural research, ERS has been mentioned as a possible source of bias in measurement (Lau, 2007; Greenleaf, 1992).

Although measures designed solely to measure ERS exist, they are not often used in studies that report ERS. For instance, the most prominent scale designed specifically for ERS is the 16 item scale by Greenleaf (1992). This scale consists of uncorrelated items such that the ERS measure is not correlated with item content. Although a common scale used by all researchers is ideal, especially when comparing ERS across studies, the additional cost and time associated with using an additional measure in a study that is not specifically focused on ERS often prohibits its application (Morren, Gelissen, & Vermunt, 2011; de Jong, Steenkamp, Fox, & Baumgartner, 2008).

Research Hypotheses

Our hypotheses are guided by past research in the area of ERS and theories that these researchers have used to interpret their findings. The studies referenced in this section were fully or in part devoted to studying ERS. This meta-analysis includes many studies that did not focus exclusively on ERS but reported enough information for inclusion in this study. This is a strength of this study, and meta-analysis in general, that it provides more concrete evidence to support its findings, by including many more data points than would be covered in a review focusing on studies specific to the topic at hand.

Our first hypothesis pertains to race. Past research has shown Blacks, on average, are more likely to engage in extreme responding than Whites (Bachman & O'Malley, 1984b; Berg & Collier, 1953). Hispanic respondents have been shown to engage in higher degrees of ERS than Whites (Clarke, 2000; Hui & Triandis, 1989) and Asian-Americans have been shown to engage in less ERS than White-Americans (Grandy, 1998). Bachman et al. (2010) found that Blacks were most likely to engage in ERS followed by Hispanics, with Whites and Asians exhibiting the lowest levels of ERS. Theories attempting to explain these findings were discussed earlier, but they tend to center on how different races tend to express themselves and the degree of collectivism of cultures. Thus, we offer:

Hypothesis 1: Black and Hispanics show more ERS than Whites and Asians.

Research concerning sex differences in ERS in this area is contradictory but the magnitude of sex differences tend to be small. Some studies report no difference for gender with respect to ERS (i.e. Greenleaf, 1992; Zuckerman et al., 1965). Others show females engaging in more ERS than males (Adams & Berg, 1961). In Lau and Howard's (2005) qualitative review, it was concluded that the preponderance of evidence on ERS suggest that females engage in more ERS than males. Thus, we offer:

Hypothesis 2: Females show more ERS than males.

Nationality differences in ERS may impact cross-cultural research. It has been argued that more collectivist societies are less likely to engage in ERS than more individualistic societies (Bettencourt & Dorr, 1997). The United States (U.S.) is considered one of the most individualistic societies. Thus, one might expect U.S. respondents to engage in higher levels of ERS than those in other countries, particularly collectivist nations. Prior research seems to support this argument regarding cultural differences in terms of response bias (e.g., Bachman & O'Malley, 1984a, 1984b; Chen, Lee, & Stevenson, 1995; Hui & Triandis, 1989; Marin, Gamba, & Marin, 1992; Triandis, 1995). For example, Zax and Takahashi (1967) found US respondents to exhibit higher ERS than Japanese. Stening and Everett (1984) indicated that people from collectivistic cultures tend to be modest in their response. Marshall and Lee (1998) found that low level of individualism relates to low levels of ERS. Chen et al. (1995) found Chinese, Canadian, and Japanese respondents to engage in less ERS than U.S. respondents, and Chun, Campbell, and Yoo (1974) found Korean respondents to engage in less ERS than those in the U.S. Despite the frequent discussion of ERS in cross-cultural studies, we located relatively few effect sizes. Given the large number of countries and the relatively few studies examining pair-wise country differences, we are hesitant to form hypotheses. We do, however, analyze the country comparisons for which we found data.

Most studies report a negative relationship between intelligence and ERS (Bregelmann, 1959; Das & Dutta, 1969; Hamilton, 1968). Meisenberg and Williams (2008) used data aggregated such that the unit of analysis was a country. Combining a country-level ERS measure derived from one data set with a country-level intelligence measure from a different data set, Meisenberg and Williams concluded that high ERS is associated with lower intelligence. Thus, we offer:

Hypothesis 3: Intelligence is negatively correlated with ERS.

There is some research on the relationship between acquiescence and ERS. For instance, Baumgartner and Steenkamp (2001) examined five forms of stylistic responding and found that the correlation between ARS (i.e., agreement tendency) and ERS is .59 and that the correlation between DARS (i.e., disagreement tendency) and ERS is .41. van Herk, Poortinga, and Verhallen (2004) pointed out a relationship between acquiescence and ERS. van Herk and her colleagues found that the number of response categories of acquiescence scale can influence the correlation between acquiescence and ERS. They reported a correlation of .23 between acquiescence and ERS is reported when two response categories of acquiescence scale are included. Nevertheless, when only one response category of acquiescence scale is incorporated, the correlation between acquiescence and ERS dramatically increases to .45. In either case, the positive correlation between acquiescence and ERS is non-trivial. At the country level, Meisenberg and Williams (2008) reported a relationship between both education and intelligence with acquiescence. Consistent with above, we advance the following hypothesis.

Hypothesis 4: Acquiescence is positively correlated with ERS.

Not all ERS-relevant studies report standardized mean differences or other correlations for ERS. However, some of these studies report mean levels of age and years of education for the same sample as well as the average across respondents for the percent of items in which extreme responses were given. Other studies report average extreme responses and the number of points in the scale (e.g., a Likert with 5 response options would have 5 “points”) and/or the number of items in the scale from which the ERS score was derived.

The meaning of average across respondents for the percent of responses to items that were judged extreme can be misunderstood. To illustrate the calculation of the measure, consider a hypothetical sample with two people, Manny and Moe. If Manny was judged to respond extremely to 30 percent of the items and Moe was judged to respond extremely to 40 percent of the items, the sample would have an average of 35 percent extreme responses. Although one can calculate this ERS percentage in the same way across studies, the criteria used to judge an item response as extreme varied across studies. For example, with a 9-point scale, one study may consider a score of 1 or 9 to be an extreme response. Another study, also with a 9 point scale, might consider responses of 1, 2, 8, or 9 to be an extreme response. Thus, the ERS percentages are not strictly comparable across studies. Yet all the measures reflect tendencies to engage in ERS and one can calculate a vector correlation (Jensen, 1998; Hunter & Schmidt refer to vector correlations as study characteristics correlations) for such data. One vector would be the ERS percentage in the sample. The second vector would be the correlate variable (e.g., age). One would weight the data contributing to the vector correlation by sample size. Because of the cross-study variability in how a response to an item is judged extreme, the resulting correlation is likely to be an underestimate of the correlation based on data in which extreme response was calculated identically across all samples.

A vector correlation analysis can shed light on the relationship between ERS and age. As reviewed earlier in the discussion of demographic correlates, the results are very mixed, and as such we do not offer a hypothesis for age but do present an analysis of available data.

The literature for education and ERS is also contradictory. Marin et al. (1992) and Greenleaf (1992) reported that higher education is associated with less ERS. However, Bachman and O'Malley (1984b) found the correlation between ERS and education to be near zero. The Bachman and O'Malley analyses

are longitudinal school-based samples where the initial variance in education is relatively low. Thus, we argue that analyses of education and ERS should consider the variance of education in the samples. Given the conflicting evidence on level of education and ERS, we offer no hypothesis for education but present an analysis of available data.

A vector correlation analysis can also provide knowledge of the relationship with number of points in the scale and number of items in the scale. We could locate no literature on these correlates and thus offer no hypotheses. As is the case with age and education, we present analyses on available data.

METHODS

Literature Search

An initial search for data was conducted using *Google Scholar*, Harzing's *Publish or Perish*, *PsycNET* (e.g., *PsycINFO*), *Web of Science* (e.g., *Social Sciences Citation Index*), and *EBSCOhost* which includes databases such as *Academic Search Complete*, *Business Source Complete*, *CINAHL plus*, *Education Research Complete*, *ERIC*, *Health and Psychosocial Instruments*, *Psychology and Behavioral Sciences Collection*, and *Teacher Reference Center*. The keywords used for these searches were "extreme response style", "extreme response", "response style", and "extreme responder." *ProQuest Dissertations and Theses* was used to identify relevant unpublished dissertations and theses. Leaders in the area of response style research were contacted for leads on additional articles and provided access to unpublished data on the topic of extreme response style. Further, reference sections of identified articles were reviewed for additional articles. This search resulted in 174 journal articles, dissertations, theses, and conference papers published from 1953 to currently in press articles. The starting date of 1953 was chosen because it is the data of the earliest known study on ERS (Berg, 1953).

Decision Rules

To be included in this review, a study needed to measure ERS on a Likert scale or end points of a semantic differential scale where respondents present their answers using adjectives on a numbered or lettered scale (i.e., good to bad, dirty to clean, awful to nice, etc.).

Data were obtained for all available extreme response mean scores, frequency of extreme response, percentage of extreme response, and correlates of extreme response. The most commonly reported measure of extreme response was frequency, followed closely by mean extreme response.

All available demographic and individual difference information was coded to permit an analysis of their relationship with extreme response style. When possible, demographic information, such as age, gender, race, and country of origin were recorded. Ethnicity was coded as Asian, Caucasian, Hispanic, African American, mixed, or other. These ethnicity groups were chosen based on the frequency of reporting in the studies. One analysis compared Caucasian scores to those of "Minorities," in this instance all non-Caucasian groups were included in the group "Minority." Samples from many different nations were obtained; these samples were coded by the country of the respondents (i.e., a study in Japan with all Japanese respondents would be coded as Japanese). For one analysis, the countries of Japan and Taiwan (Chen et al., 1995) were combined into one Asian group for analysis.

Meta-Analytic Procedures

All meta-analyses were performed using the Comprehensive Meta-Analysis (CMA) software developed by Borenstein, Hedges, Higgins, and Rothstein (2005). Such analyses are different but computationally similar to "bare bones" meta-analysis in psychometric meta-analysis (Hunter & Schmidt, 2004). Vector correlations were calculated in SPSS. We could locate very little data on the reliability of measures. Thus, psychometric meta-analysis with corrections for measurement error could not be conducted. The references for the studies included in the present meta-analysis were marked with an asterisk in the reference section. The tables that contain the codes/data to the present meta-analysis are available upon request.

RESULTS

Table 1 presents meta-analysis results that address race, sex, and nationality relationships with ERS. The table also presents meta-analyses of correlations between both intelligence and acquiescence with ERS. Table 2 presents vector correlation analyses addressing the relationship between ERS and age, education, number of points in a scale and number of items in a scale.

TABLE 1
META-ANALYSIS OF STANDARDIZED MEAN DIFFERENCES BETWEEN GROUPS ON
ERS AND CORRELATIONS OF INTELLIGENCE AND ACQUIESCENCE WITH ERS

<u>Demographic Mean Differences</u>							
Distribution	# Samples	# Studies	<i>N</i>	<i>d</i>	Confidence Interval	<i>I</i> ²	
Race							
White-Black	9	3	232,327	-.254	-.505 to -.002	99.789	
White-Hispanic	8	2	226,986	-.089	-.130 to -.048	87.063	
White-Asian	6	1	209,114	.158	.137 to .179	-	
White-Minority (multiple minorities)	10	3	268,571	-.135	-.173 to -.097	91.365	
Sex							
Female-Male	19	6	287,510	.090	.055 to .125	93.012	
<u>Country Mean Differences</u>							
Distribution	# Samples	# Studies	<i>N</i>	<i>d</i>	Confidence Interval	<i>I</i> ²	
US- Japan	1	1	3,118	.234	.157 to .311	-	
US-Asian (multiple Asian groups)	1	1	4,475	.271	.212 to .329	-	
US-Mexico	1	1	504	-.148	-.323 to .027	-	
US-Canadian	1	1	2,861	.267	.181 to .353	-	
US-Australian	1	1	504	-.128	-.047 to .303	-	
US-French	1	1	504	.175	.00 to .350	-	
<u>Correlations with ERS</u>							
Distribution	# Samples	# Studies	<i>N</i>	<i>r</i>	Confidence Interval	<i>I</i> ²	
Intelligence	2	2	231	-.260	-.377 to -.135	0.000	
Acquiescence	6	6	110,668	.266	.078 to .436	99.817	

Note: A negative *d* indicates the first group has a less extreme response than the second group; a positive *d* indicates that first group has more extreme responding than the second group. For example, the White-Black *d* of -.254 indicates that Whites have a mean of extreme responding that is about one-fourth of a standard deviation below the Black mean. A Female-Male *d* of .09 indicates that females have slightly more extreme responding than males.

TABLE 2
VECTOR CORRELATIONS ANALYSES

Distribution	# Samples	# Studies	<i>N</i>	Vector <i>r</i>	Confidence Interval
Age (linear relationship)	15	3	16,678	-.649	-.657 to -.641
Age (non-linear relationship)	15	3	16,678	.772	.776 to .778
Education	16	3	16,543	-.009	-.024 to .006
Education (Low Variance)	9	2	15,736	-.012	-.028 to .004
Education (High Variance)	7	2	807	-.477	-.529 to -.422
Number of Points in Scale	15	6	225,728	.263	.260 to .266
Number of Items in Scale	15	6	10,191	.192	.174 to .210

Hypothesis 1 concerned race differences. Nine samples with a total sample size (*N*) of 232,327 examined White-Black ERS differences yielding a *d* of -.254 (see Table 1 for this and other mean racial comparisons). Thus, Whites engage in less ERS than Blacks. The *d* is interpreted in standard deviation units. Expressed in another way, Blacks are about one-fourth standard deviation higher in ERS than Whites. With respect to White-Hispanic differences, data were available on eight samples (*N* = 226,986). Hispanics engaged in ERS slightly more often than Whites (*d* = -.089). The data are supportive of Hypothesis 1 in that Blacks and Hispanics showed more ERS than Whites. The confidence intervals for the *d* do not include zero so that differences between Whites and Blacks and the differences between Whites and Hispanics are statistically significant. Readers can view the confidence intervals for all analyses to draw inferences regarding statistical significance.

We located no data comparing Asians with Blacks or Asians with Hispanics, but we did locate 6 samples (*N* = 209,114) comparing Whites with Asians. In the data, Whites engaged in more ERS than Asians.

There were 10 samples (*N* = 268,571) contrasting Whites with Minorities in which the minority group consisted of multiple non-White races. Whites showed less ERS than the racially-heterogeneous minorities. Given that Asians show less ERS than Whites, but Blacks and Hispanics show more ERS than Whites, the analysis of a racially-mixed minority group does not have a clear interpretation. If the majority of a racially-mixed minority group is Black, then these results are consistent with the White-Black analysis.

Hypothesis 2 postulated that females will show more ERS than males. The comparison by sex is based on 19 samples (*N* = 287,501). Females had slightly more ERS than males (*d* = .09).

Hypotheses 3 offered that intelligence is negatively related to ERS and is supported by the correlation of -.260 reported in Table 1. The analysis is based only on 2 samples (*N* = 231).

The final analysis shown in Table 1 addresses hypothesis 4 and shows the correlation between acquiescence and ERS. The results are based on five samples (*N* = 101,947) and yields a correlation of .204.

Table 2 presents vector correlation analyses for several variables. In these vector correlation analyses, the unit of analysis is the sample. Thus, the sample size for the vector correlation is the number of samples and the confidence interval is based on a standard error of the correlation which is a function of the number of samples.

Age yields a vector correlation of -.649 suggesting that ERS declines with age. Because previous authors (Bachman et al., 2010; Greenleaf, 1992) have suggested an inverted U shaped relationship between age and ERS, we examined curvilinearity between age and ERS. To examine the possibility of a curvilinear relationship with one bend, we used a stepwise regression entering age in step one and age-squared in the second step. A departure from linearity would be supported by an increase in the *R*² from step 1 to step 2. The *R*² increased from .649 to .772 suggesting a curvilinear relationship. The non-linearity is concave with ERS increasing with age until the early 20's and then declining with age.

With respect to other vector correlation analyses, years of education was unrelated to ERS (Vector $r = -.009$). However, the variance of education in the sample was a strong moderator. For samples with low variance in education, the vector correlation with ERS was very small ($-.012$). However, for samples with larger variance, the vector correlation was much larger ($-.477$). Thus, with educational level heterogeneity in a sample, one can expect less educated samples to show more ERS.

ERS increased as the number of points in a scale increased (.263). Perhaps increasing the range of rating scale points enhances the reliability of the ERS measurement. ERS also increased with the number of items in the scale on which ERS was measured (.192). As with the number of points on a scale, increasing the number of items likely increases the reliability of ERS measurement.

DISCUSSION

This paper sought to examine relationships between ERS and other variables. The following discussion provides conclusions on what can be drawn from this study and identifies current gaps in the ERS literature.

Our findings support mean racial differences in ERS with the largest being the White-Black difference. Blacks, on average, engage in more ERS than Whites ($d = .254$). Hispanics, on average, engage in slightly more ERS than Whites ($d = -.089$). Asians, on average, engage in slightly less ERS than Whites ($d = .158$). The ERS sex differences were also small with females engaged in ERS slightly more so than males ($d = .090$).

Our conclusions about nationality differences must be tentative because we located only one effect size for each country comparison. U.S. residents, on average, displayed more ERS than Asians, Canadians and French and less ERS than Mexicans and Australians. Given that cross-cultural researchers often speculate about ERS differences, and their effects, research on cross-cultural ERS differences is encouraged. Researchers should provide sufficient data on the difference (i.e., means and standard deviations) such that a standardized mean difference can be calculated. Consistent with suggestions by Meisenberg and Williams (2008) for data at the country level, we conclude that intelligence is negatively correlated with ERS ($r = -.260$) at the level of the individual but offer that conclusion as tentative given that there were only two small samples ($N = 231$).

We suggest that that the intelligence correlates, the education correlates, and the race correlates may be related. The ordering of the races by ERS from least to most is Asians, Whites, Hispanics, and Blacks. This corresponds to the likely ordering of mean intelligence by race (Rushton, 2000). In Table 1, we report a negative correlation between intelligence and ERS. Thus, we speculate that the mean racial differences in ERS may be, at least in part, a function of intelligence. For samples with large variance in educational level, the vector correlation with ERS was moderate ($-.477$; $N = 807$). Because educational level correlates with intelligence and because race correlates with both educational level and intelligence, we suggest the mean racial differences in ERS may be a function of intelligence and/or educational differences. Future research is warranted in this area.

Another topic for greater exploration is the positive correlation between ERS and acquiescence. Our meta-analytic result yields a correlation of .266 between ERS and acquiescence, which is compatible with the findings from Baumgartner and Steenkamp (2001) and van Herk et al. (2004) and is consistent with country-level effects reported by Meisenberg and Williams (2008). Our finding thus corroborates the argument that a modest correlation between ERS and acquiescence may be commonly expected.

The vector correlations are also informative. The vector correlation analyses for age with ERS suggest a strong relationship. The linear relationship (Vector $r = -.649$) suggest that ERS drops sharply with age. However there is also evidence for a curvilinear effect such that ERS increases with age until the early 20's and then declines. The vector correlation with education is near zero which is at odds with the finding for intelligence. However, some of the studies reported limited variance on education and conclusions concerning education and ERS must await additional research. The vector correlations with number of points in the scale and number of items in the scale are both positive. This suggests that these

correlations are a function of increased reliability of measure in ERS as the number of scale points increase and the number of items in the scale used to measure ERS increases.

LIMITATION AND FUTURE RESEARCH

There are some important limitations to our findings. First, the interpretation of the means of the distributions used in the meta-analysis is complicated by the large I^2 values. The statistic I^2 is the percentage of variance that is not attributed to sampling error (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003). A large I^2 typically indicates the presence of a few large magnitude moderators or many small magnitude moderators. Thus, our meta-analytic means are from distributions with substantial true (population) variability. For some moderator subgroups, the d can be expected to be larger than reported here and in other subgroups the d could be smaller. This suggests the need for large sample analyses where the prediction of ERS can be addressed in a manner that considers multiple moderators simultaneously.

Thus large sample datasets would allow the testing of some of the suggested methods for dealing with ERS. One is Latent Class Factor Analysis (LCFA) (Moors, 2003; Morren et al., 2011). This method controls for ERS as it would any other method factor, such as reverse scoring, by creating a model that contains all latent variables of interest along with an ERS method factor. Here the effects of ERS are controlled by the introduction of the method factor which provides an indication of the strength of effect of ERS and the interpretation of the relationships of interest with the influence of ERS controlled.

The second limitation is the relatively small number of samples. Even though our sample sizes tend to be very large, any given sample has a set of unique characteristics. Some of these may influence the relationships with ERS. These sample characteristics likely are responsible for the heterogeneity of the effect size distributions.

The third limitation is related to the limitations on inferences that can be drawn from vector correlations. Conclusions based on vector correlations reflect study level relationships. Generalizations to individual data should be made cautiously.

CONCLUSION

This paper has investigated relationship between ERS and several variables. We concur with past research in the conclusion that ERS presents a measurement challenge that is not widely known or addressed in much research involving Likert or similar scales. We present the first quantitative review of this literature to examine the correlates of ERS. This study found that there are indeed differences in how certain groups of individuals tend to use the ends of response scales. Our findings reveal that White, Black, Hispanic, and Asian groups have very different tendencies with respect to ERS. This leads us to strongly suggest that researchers address the nuisance effects of ERS whenever engaging in cross-cultural research. Further, a small magnitude difference in gender was found. Age and intelligence were found to have negative relationship with ERS. Age also has a non-linear relationship with ERS. And finally, reducing the number of items and points in scales were offered as avenues for future researchers to investigate in mitigating the effects of ERS on cross-cultural research.

REFERENCES

- Adams, H.E., & Berg, I.A. (1961). Affective tone of test option choice as a deviant response. *Psychological Reports, 8*, 79-85.
- *Albaum, G., Roster, C., Yu, J.H., & Rogers, R.D. (2007). Simple rating scale formats. *International Journal of Market Research, 49*, 633-650.
- Bachman, J.G., & O'Malley, P.M. (1984a). Black-white differences in self-esteem: Are they affected by response styles? *American Journal of Sociology, 90*, 624-639.

- Bachman, J.G., & O'Malley, P.M. (1984b). Yea-saying, nay-saying, and going to extremes: Black-White differences in response style. *Public Opinion Quarterly*, 48, 491-509.
- *Bachman, J.G., O'Malley, P.M., & Freedman-Doan, P. (2010). *Response styles revisited: Racial/ethnic and gender differences in extreme responding* (Monitoring the Future Occasional Paper No. 72). Ann Arbor, MI: Institute of Social Research. Available from <http://www.monitoringthefuture.org/>
- Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Berg, I.A. (1953). The reliability of extreme position response sets in two tests. *Journal of Psychology*, 36, 3-9.
- Bettencourt, B. A., & Dorr, N. (1997). Collective self-esteem as a mediator of the relationship between allocentrism and subjective well-being. *Personality and Social Psychology Bulletin*, 23, 955-964.
- *Berg, I. A., & Collier, J. S. (1953). Personality and group differences in extreme response sets. *Educational and Psychological Measurement*, 13, 164-169.
- Bockenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, 17, 665-678.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.
- Brengelmann, J.C. (1959). Differences in questionnaire responses between English and German nations. *Acta Psychologica*, 16, 339-355.
- *Chen, C., Lee, S., Stevenson, H.W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6, 170-175.
- Chun, K., Campbell, J.B., & Yoo, J.H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology*, 5, 465-480.
- Clarke, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavioral and Personality*, 15, 137-152.
- *Clarke, I. (2001). Extreme response style in cross-cultural research. *International Marketing Review*, 18, 301-324.
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 60, 151-174.
- Cronbach, L.J. (1946). Response set and test validity. *Educational and Psychological Measurement*, 6, 475-494.
- Cronbach, L.J. (1950). Further evidence of response set test design. *Educational and Psychological Measurement*, 10, 3-31.
- *Das, J.P., & Dutta, T. (1969). Some correlates of extreme response set. *Acta Psychologica*, 29, 85-92.
- de Jong, M.G., Steenkamp, J.E.M., Fox, J., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, 45, 104-114.
- *Dick, N. (1976). *The extreme response style as a function of psychotherapy and meaning fullness: A test of the deviation hypothesis* (Unpublished master's thesis). University of Miami, Florida.
- Grandy, J. (1998). Response bias in a survey of Asian-American and White science and engineering students. *Journal of Women and Minorities in Science and Engineering*, 4, 1-13.
- *Greenleaf, E.A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351.
- *Gruber, R.E. (1979). *The impact of response style and response set on large scale survey research involving activities, interests and opinions (AIO) variables*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 7904082)
- *Hamilton, D. L. (1965). *An analysis of extreme response set as related to personality*. (Unpublished master's thesis). University of Richmond, Virginia.
- Hamilton, D.L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, 69, 192-203.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.

- Higgins, J., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal*, *327*, 557-560.
- *Hui, C.H., & Triandis, H.C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, *20*, 296-309.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. (2nd ed.). Newbury Park, CA: Sage.
- Iwawaki, S., Zax, M., & Mitsuoka, S. (1969). Extremity of response among Japanese and American children. *Journal of Social Psychology*, *79*, 257-259.
- *Hurley, J.R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology*, *132*, 201-210.
- Jackson, D.N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, *55*, 243-252.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger
- *Johnson, T., Kulesa, P., Chou, Y.I., Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*, 264-277.
- Lau, M.Y. (2007). *Extreme response style: An empirical investigation of the effects of scale response format and fatigue* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3299156)
- Lau, M. Y., & Howard, G. S. (2005). *Extreme response style: A review of the evidence, assessment and potential solutions to the problem*. Poster presented at the Division 17 Student Poster Session at the American Psychological Association Annual Convention in Washington, DC.
- Marin, G., Gamba, R.J., & Marin, B.V. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross-Cultural Psychology*, *23*, 498-509.
- Marshall, R., & Lee, C. (1998). A cross-cultural, between-gender study of extreme response style. In B. G. Englis, & A. Olofsson (Eds.), *European advances in consumer research* (pp. 90-95). Provo, UT: Association for Consumer Research.
- McDaniel, M.A., Pstotka, J., Legree, P.J., Yost, A.P., & Weekley, J.A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology*, *96*, 327-336.
- Moors, G. (2003). Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality and Quantity*, *37*, 277-302.
- *Meisenberg, G., & Williams, A. (2008). Are acquiescence and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*, 1539-1550.
- Moors, G. (2004). Facts and artifacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *European Sociological Review*, *20*, 303-596.
- Morren, M., Gelissen, J.P.T., & Vermunt, J.K. (2011). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. *Sociological Methodology*, *41*, 13-26.
- Naemi, B.D., Beal, D.J., & Payne, S.C. (2009). Personality predictors of extreme response style. *Journal of Personality*, *77*, 261-286.
- Neuringer, C. (1961). Dichotomous evaluations in suicidal individuals. *Journal of Consulting Psychology*, *25*, 445-449.
- Novick, M.R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*, 1-18.
- *Norman, R.P. (1969). Extreme response tendency as a function of emotional adjustment and stimulus ambiguity. *Journal of Consulting and Clinical Psychology*, *33*, 406-410.
- Oskamp, S. (1977). Methods of studying social behavior. In L.S. Wrightman (Ed.), *Social Psychology*. Monterey, California: Brooks and Cole.

- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wright (Eds.), *Measurement of personality and social psychology attitudes* (pp. 17-59). New York: Academic Press.
- Rorer, L.G. (1965). The great response-style myth. *Psychological Bulletin*, *63*, 129-156.
- Rushton, J.P. (2000). *Race, evolution, and behavior: A life history perspective (3rd Edition)*. Charles Darwin Research Institute: Port Huron, MI.
- Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology*, *122*, 151-333.
- Triandis, H.C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, *35*, 346 – 360.
- *Van Ijzendoorn, M.H. (1984). Answers without questions: A note on response style in questionnaires. *Perceptual and Motor Skills*, *59*, 827-831.
- *Watkins, M. L. (1992). The implications of extreme response style (ERS) for cross-cultural and comparative research in South Africa. *Journal of Industrial Psychology*, *18*, 13-19.
- *Weech-Maldonado, R., Elliott, M.N., Oluwole, A., Schiller, K.C., & Hays, R.D. (2008). Survey response style and differential use of CAHPS rating scales by Hispanics. *Medical Care*, *46*, 963-968.
- Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2015). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 1-13.
- *Wilkinson, A. E. (1970). Relationship between measures of intellectual functioning and extreme response style. *The Journal of Social Psychology*, *81*, 271 – 272.
- *Zax, M., Gardiner, D. H., & Lowy, D. G. (1964). Extreme response tendency as a function of emotional adjustment. *Journal of Abnormal and Social Psychology*, *69*, 654-657.
- Zax, M., & Takahashi, S. (1967). Cultural influences on response style: Comparisons of Japanese and American college students. *The Journal of Social Psychology*, *71*, 3-10.
- Zuckerman, M., Oppenheimer, C., & Gershowitz, D. (1965). Acquiescence and extreme sets of actors and teachers. *Psychological Reports*, *16*, 168-170.

*The papers meta-analyzed in this study are marked with an asterisk.