Katarzyna Kinga Kowalczyk, Agnieszka Mulak, Grzeorz Król

Extreme Response Style in Correlational Research

Problemy Zarządzania 14/2 (2), 61-82

2016

Problemy Zarządzania, vol. 14, nr 2 (60), t. 2: 61–82
ISSN 1644-9584, © Wydział Zarządzania UW
DOI 10.7172/1644-9584.60.4

Extreme Response Style in Correlational Research

Submitted: 15.06.16 | Accepted: 18.08.16

Katarzyna Kinga Kowalczyk*, **Agnieszka Mulak****, **Grzegorz Król*****

Extreme Response Style, defined as the tendency to consistently use (or avoid) the extreme ends of response scales in questionnaires with a multiple response category format (Clark, 2000), is one of the most studied response styles. The reason for this is its potential to distort survey results, especially in cross-cultural research (Cheung & Rensvold, 2000). Furthermore, the relationships between ERS and most frequently analyzed socio-demographic variables (such as gender, age, education) are inconclusive. We replicated a number of well-established effects, using data from the World Value Survey, and expanded these analyses by including an ERS index, to determine its possible effect on the original results. We found no significant effects of ERS in any of the replicated studies, as long as the ERS index is uncorrelated with the dependent variable. We followed with a theoretical model and a recount of simulation results in an attempt to define boundaries for the phenomenon of ERS significantly affecting research conclusions. We argue that even for interval-level correlation measures, e.g. as used in the ordinary least squares regression analysis, the effect on real data is negligible.

Keywords: response style, extreme response style, survey research.

Extreme Response Style (Skrajny Styl Odpowiadania) w badaniach korelacyjnych

Nadesłany: 15.06.16 | Zaakceptowany do druku: 18.08.16

Extreme Response Style (Skrajny Styl Odpowiadania), definiowany jako skłonność do konsekwentnego korzystania (albo unikania) z krańcowych odpowiedzi w kwestionariuszach (Clark, 2000), jest jednym z najczęściej badanych stylów odpowiadania. Uważa się, że może on zniekształcać wyniki analiz, szczególnie w badaniach międzykulturowych (Cheung i Rensvold, 2000). Wyniki badań nad związkiem ERS z innymi zmiennymi (np. płcią, wiekiem) również są niejednoznaczne. W celu sprawdzenia wpływu ERS, wykorzystując dane z World Value Survey, zreplikowaliśmy kilka analiz, dołączając do nich wskaźnik ERS. W przypadku, gdy wskaźnik ERS nie był skorelowany ze zmienną zależną, nie wpływał on na wyniki. W następnym kroku przeprowadziliśmy symulacje mające na celu określenie sytuacji, w których ERS mógłby istotnie wpływać na relacje pomiędzy zmiennymi. Uważamy, że nawet w przypadku korelacji mierzonych na skalach przedziałowych wykorzystywanych np. w analizie regresji metodą najmniejszych kwadratów, efekt ten jest zaniedbywalny.

Słowa kluczowe: style odpowiedzi, skrajny styl odpowiedzi, badania sondażowe.

JEL: C18

* **Katarzyna Kinga Kowalczyk** – mgr, Department of Managerial Psychology and Sociology, University of Warsaw.

** **Agnieszka Mulak** – mgr, Institute of Social Studies, University of Warsaw.

*** **Grzegorz Król** – dr, Department of Managerial Psychology and Sociology, University of Warsaw.

Correspondence address: University of Warsaw, 1/3 Szturmowa Street, 02-678 Warsaw, Poland; e-mail: gkrol@wz.uw.edu.pl; kkowalczyk@wz.uw.edu.pl; ak.mulak@gmail.com.

Ministry of Science and Higher Education
Republic of Poland

The creation of the English-language version of these publications is financed in the framework of contract No. 768/P-DUN/2016 by the Ministry of Science and Higher Education committed to activities aimed at the promotion of education.

1. Introduction

It is widely assumed that when people fill out a questionnaire, their answers are based entirely on the substantive meaning of items. However, their responses may also be influenced by various content-independent factors, including the rating scale associated with an item (Baumgartner & Steenkamp, 2001). In most questionnaires, respondents mark their opinion on a response scale provided by the researcher, which in turn requires that they map their private response scales onto the provided one. This transposition determines their response style, which can be universal, i.e. relatively independent of the object of assessment (Wieczorkowska, 1993; WWK, 2014). Response style may be defined as the tendency to respond consistently to questionnaire items on any basis other than what the items were specifically designed to measure (Harzing et al., 2011). This effect is particularly problematic in cross-cultural comparisons and may lead to drawing erroneous conclusions about particular respondent groups differing with regard to the phenomenon under investigation, whereas the groups only differ in terms of their response styles. It is said that a response style can inflate or deflate correlations between respondents' scores on various scales and influence conclusions concerning the relationship between variables (Baumgartner & Steenkamp, 2001).

Differences in response styles usually manifest themselves through the tendency to employ one of the following behavior patterns:

1. avoiding expressing one's opinion, by frequently choosing a "Don't know" response (DK);
2. favoring the extremes of the scale (Extreme Response Style, ERS) over the middle (Midpoint Response Style, MRS) responses of the scale;
3. agreeing (Acquiescence Response Style, ARS) or disagreeing (Disacquiescence Response Style, DRS) with an item regardless of the content.

Our topic of interest is Extreme Response Style defined as the tendency to consistently use the extreme ends of response scales in questionnaires with a multiple response category format (Clark, 2000). A review of literature revealed several interesting questions regarding the role of ERS: how to measure ERS, what impact it may have on cross-cultural comparisons, what are the relationships between ERS and other variables (sex, age, education, personality traits). Answers to these questions still remain ambiguous – and differ between various authors (see: Greenleaf, 1992; Cheung & Rensvold, 2000). In this article we focus on the following issues:

1. Does ERS constitute sufficiently important influence to be controlled for "by default" in the way basic demographic factors usually are in cross-cultural research?
2. Does ERS influence correlational analysis?

Our research aims to discern under what circumstances ERS tends to be insignificant and can be excluded from researchers' considerations.

 DOI 10.7172/1644-9584.60.4

This article is organized as follows. In section 2 we summarize the ways of analyzing ERS, and common problems with ERS measurements. In section 3 we give an overview of ERS research up to date. In section 4 we report the results of replications of three published studies, extending the analyses to include ERS control. We follow with a theoretical model in section 5 and recount simulation results in order to define boundaries in which the phenomenon of ERS might significantly affect research conclusions. In section 6 we conclude with future directions of work on ERS.

2. ERS – approaches

Extreme Response Style (ERS) is considered by some researchers as an important factor to take into account in multivariate models, whether they involve just demographic measures, or soft, theoretical constructs. Two different approaches can be considered:

1. ERS may be analyzed as a mediator or moderator of the relationship between DV and IVs in the model.
2. ERS may be considered as one of the standard IVs in a model. It can be considered either as a controlled variable, or included as an important part of the model.

Within the first approach, one strives to understand how the formal response style influences relationships between the DV in the model and other IVs. This approach is useful e.g. when the model consists of demographic IVs only, and when we suspect that the ERS might help us better understand the relationship between IVs and the DV, by increasing the amount of variance explained by DVs in groups defined by different levels of ERS, or modifying the slope of relationship between DV and IVs, depending on the level of ERS.

The second approach is applicable when one considers ERS as yet another influencing variable that needs to be controlled for – that is, ERS is treated just like other independent variables. In many models demographic IVs – like gender, education, age, socio-economic status etc. – are included as controlled variables. In an ideal experimental situation, the value of controlled variable remains constant. In many real-case scenarios of research, we can only use statistical methods to control for their effect. ERS can be treated in the same way, and included as a controlled variable into the model, prior to adding the key variables of interest. The effect of ERS is thus taken into account at a statistical level.

ERS effect can also be a part of the tested hypothesis. In this case it is included in the model not merely as a controlled factor, but as an object of interest for the researcher.

In such a situation a question arises: is it appropriate to research how ERS "influences" demographic IVs? ERS cannot causally influence any of the standard demographic variables. However, it can be potentially influ-

enced by things like gender, age or education. It is then valid to consider ERS as a moderator variable (for interaction effects) in models involving only socio-demographic IVs, if one suspects that some part of the variance of the DV might be explained by a formal style of answering.

2.1. ERS measurement problems

There is a broad discussion about operationalizing ERS. Two main approaches are sum scores and methods that are based on statistical modeling of response biases. Modeling approaches include IRT, latent class, confirmatory factor analysis (Kieruj, Moors, 2011). One of the important issues linked with measuring the ERS effect is that it is easily confounded with the content effect. While items for a single scale or subject area might seem easier to use, by doing so in ERS study a researcher risks that the potential, content independent ERS effect might be confounded with a specific effect of the subject. The solution to this problem is balancing scales or ensuring maximum heterogeneity of the items used to measure the response (Beuckelaer, Weijters & Rutten, 2009). One of the proposed methods helping to overcome this problem is an approach known as *Representative Indicators for Response Styles* (RIRS) which removes the formerly described limitation. It requires drawing a random sample of survey items from a range of available multi-item scales, and using this subsample to determine the ERS score (De Beucklear & WeijtErs, 2010).

Another method worth mentioning is the New IRT Model for Measuring ERS proposed by De Jong et al. (2008). It allows marketing researchers to construct an ERS measure based on substantively correlated items and eliminates the need for a dedicated ERS scale. This method accommodates the possibility that an item's usefulness differs across groups. Therefore it allows various items to be differentially useful for measuring ERS.

Another problem related to ERS is the number of response categories. According to Hui & Triandis (1989, cited in Clarke, 2000) increasing the number of response categories can reduce the usage of extreme responding. In his study Clarke (2000) showed that across cultures, a 3-point scale demonstrated the highest absolute level of extreme responding (but the lowest between-group difference), with a drop of extreme responses when changing into 5-point scale, and a minimum decrease when moving from 7 to 9 points formats, and the lowest between groups with a 3-point scale. He suggested that limiting the number of response categories to three might create the least bias for cross-cultural comparisons, even with higher levels of ERS than in the other formats (Clarke, 2000, p.146). Harzing et al. (2009) showed that changing the number of response options from 5 to 7 provided a slight reduction in both response and language bias. However, Kieruj & Moors (2011) found no evidence that could lead to a suggestion regarding the optimum number of response categories in terms of ERS.

 DOI 10.7172/1644-9584.60.4

3. Studies on ERS

Extensive research has been published on this topic with main issues concerning: cross-cultural comparisons, relationship with socio-demographic and cultural variables, methods of measurement and scaling problems. Some of the results are fairly inconsistent and difficult to generalize.

3.1. Cross-cultural differences

Research on ERS delivers the most congruent results in the context of cross-cultural comparisons. Studies show that in the USA Hispanics and Afro-Americans have a consistently higher rate of extreme responses than non-Hispanics (Clarke III, 2000) whereas Philippines and Americans indicate a higher mean proportion use of extreme scale points as compared to the Chinese and Irish (Roster, Albaum & Rogers, 2006). Meta-analysis conducted by Batchelor, Miao& McDaniel (2013) showed that whites engage less in ERS than blacks, but more than Asians, and Hispanics slightly more often than whites. In their studies on representative internet samples of Italian and Japanese respondents, and using data collected in the International Social Survey Program (1998, 2008), Wieczorkowska-Wierzbińska et al., (2014) showed that the Japanese typically gave fewer extreme responses than Italians, as well as more "don't know" answers. Upon examining five groups of respondents (the Chinese in China, the Chinese in Australia, Anglo-Australians in Australia and two groups of German students in Germany), Harzing et al. (2011) found national/cultural background to have a strong impact, with respondents from collectivist countries displaying a tendency towards less extreme responses.

3.2. Demographic variables

Gender

The results of studies on the relationship between ERS and gender are inconsistent. Some findings indicate that women mark extreme responses more frequently than men do (De Jong et al., 2008). A meta-analysis conducted by Batchelor, Miao & McDaniel (2013) has shown that on average females had slightly higher ERS than males. Some studies conclude the opposite (Marshall & Lee, 1998), whereas others show no differences between the sexes (Kieruj & Moors, 2013; Clark III, 2000). These inconsistencies may partially be explained by differences in the type of judgments employed in research questionnaires. As Crandal (1973) noticed, women tended to choose more extreme responses in studies involving affective evaluations, while men did so in the case of judgments that were affectively neutral or cognitive in nature. His studies also showed that greater ERS for women is more likely to occur in studies with positive rating scales than those utilizing negative scales.

Table 1 below summarizes the relationship between ERS and gender in the literature published in recent years. The summary shows a lack of consistency across individual studies, which should not be very surprising, taking into account different methodologies and samples used in these studies.

Authors	ERS index	Scale and Method	Results
De Jong et al. 2008 Sample: data from 26 countries on four continents (N = 12,506)	The authors used a heterogeneous set of 19 multi-item scales and two single items. The total number of items was 100.	Scale: five-point Likert items Method: IRT model	The socio-demographic variables explained approximately 2% of the Level 1 variance. Women tend to score higher on ERS than men
Marshall & Lee 1998 Sample: Student groups (N=604)	The authors used a 14-item scale drawn from another study. The original rating values of the data were transformed in order to represent the distance from the midpoint of the scale. Each respondent's ERS score was calculated as his or her averaged score over all fourteen items in the ERS scale. These individual mean scores formed a distribution, and the mean of this distribution was taken as the ERS score for a group.	The seven-point, Likert-type scale Method: A two-way analysis of variance	Women were found to display less extreme responses than men.
Kieruj & Moors 2013 Sample: a household panel (N = 8,044)	A selection of 23 items was made referring to seven personality traits which were expected to be related to ERS and ARS. The second set of 18 items was selected from a pool of attitudinal questions following the methodology suggested by Greenleaf (1992) which involves the summing of extreme responses on a set of items with low inter-item correlations.	Scales: short scales (5-, 6- and 7-point) and long scales (9-, 10- and 11-point) Method: a latent class confirmatory factor model	Gender produced non-significant correlations with ERS.

Tab.1. Relationship between ERS and gender – differences

Age

Results concerning the relationship between age and ERS are similarly ambiguous. Some studies indicate that ERS increases with age (Greenleaf, 1992; GWW 2014; Kieruj & Moors, 2012). On the contrary, Roster, Albaum & Rogers (2006) found the eldest group to be least likely to provide extreme responses in the Philippines. At the same time, they failed to observe any significant relationship between ERS and age in the US and Ireland. In yet

 DOI 10.7172/1644-9584.60.4

another study De Jong, Steenkamp, Fox & Baumgartner (2008) discovered that both the youngest and the oldest old were most likely to respond extremely. Age was also not significantly related to ERS in a study of Warnecke et al. (1997).

Education

More congruent results were reported concerning the relationship between ERS and education. Extreme responding is linked to low education levels (Meisenberg & Wiliams, 2008; Greenleaf, 1992b; Warnecke et al., 1997). However, De Jong, Steenkamp, Fox & Baumgartner (2008) found no cross-nationally generalizable effect for education.

3.3. Personality and cognitive abilities

Some studies found significant relationships between ERS and personality variables. Extreme responders are more likely to be extraverted (Austin et al., 2006), display lower tolerance for ambiguity and show a tendency towards simplistic thinking (Naemi, Bael & Payne, 2008). Furthermore, extreme responding is consistently associated with low intellectual ability (Meisenberg & Wiliams, 2006, cited in Meisenberg, Lawless, Lambert & Newton, 2006) and low IQ score (Meisenberg & Williams, 2008). In their meta-analysis, Batchelor, Miao & McDaniel (2013) found intelligence to be negatively related to ERS.

3.4. Culture characteristics

Studies on the relationship of ERS and culture characteristics are less numerous than those pertaining to the previous topics. Johnson et al. (2005) revealed that ERS is positively related to masculinity and power distance. Other cultural dimensions like national cultural individualism or uncertainty avoidance also play an important role in explaining cross-national differences in ERS (De Jong et al., 2008). Results for the various response styles in 26 countries showed relations with power distance, collectivism, uncertainty avoidance and extraversion (Harzing, 2006). However, in the above mentioned studies of Johnson et al. (2005), uncertainty avoidance and individualism were not independently associated with the extreme response style.

Based on the literature review we conclude that a) ERS can be approached at the individual level of analysis, or at an aggregate (often country/culture) level, b) relationships between ERS and most frequently analyzed variables (such as gender, age, education) are inconclusive. The reason for the lack of consistency in results mentioned may be due to the different measurement methods, including the ERS measurement method and a different number of response intervals considered (Greenleaf, 1992). It may as well stem from ERS independent characteristics of the studies or culture differences.

4. Replications: ERS in correlational studies

The main question we raised is whether the ERS effect on real data should be taken into consideration during data analysis and interpretation or whether it is negligible and can be omitted. We began investigating this topic by replicating several published data analyses and including the ERS measure into our calculations. We aimed to observe whether ERS would influence the outcomes. If the effect of ERS is negligible, results of the replications with ERS included as an additional explanatory variable, and without it, should not differ significantly.

In order to maintain relatively consistent quality of data used in our study, we undertook to replicate effects of published analyses of World Value Survey data (waves 3, administered from 1994 to 1999 and 5, administered from 2005 to 2009), conducted at both individual and country level.

4.1. Operationalizing ERS in replications

Our intention was to single out a measure that can be easily reproduced in many contexts and in various subject areas. We decided on a method well described in literature. Utilizing a measure that requires complex statistical pre-processing would severely limit its usability by other researchers. Taking that into consideration, we weighted pros and cons of each available method, and decided on the one involving the count of extreme answers (regardless of the scale length) to build ERS indices for further analyses. We created two ERS indices using the *Representative Indicators for Response Styles* (RIRS) approach, for each of the three replicated studies. For ERS index one, a random choice of 10 variables was made. For index two, additional 10 variables were randomly selected. Out of the 20 variables (re-using those from index one), for each replicated study we selected 10 variables with the lowest correlations with the dependent variable under investigation.

4.2. Study 1: Why are conservatives happier than liberals?

The exact replication of Napier & Jost's (2008) "Why conservatives are happier than liberals" analyses turned out to be a very difficult task, due to their case selection methodology. We replicated the second of three studies in the original paper. The aim of this study was to explore the relation between political orientation and subjective well-being at an individual level. The authors hypothesized that endorsing a meritocratic belief system would account for the relation between political conservatism (or right-wing orientation) and subjective well-being. The data came from the third wave of the World Values Survey. First, Napier & Jost constructed a stepwise linear regression model to predict subjective well-being in the United States. Political orientation was entered in the first step, adjustment variables were entered in the second and third steps, and endorsement of

 DOI 10.7172/1644-9584.60.4

meritocracy was entered in the fourth step. They also constructed a multilevel model, adjusting the intercept of each nation for GDP, unemployment rate, inflation rate, and the HDI.

After replicating the authors' analyses, we included two versions of the ERS index in the regression analysis. The ERS1 index (created with the random variable selection procedure) was significantly correlated with the dependent variable (subjective well-being), $r = 0.143$ ($p < 0.001$). ERS2 was constructed from a subset of randomly selected variables, choosing 10 variables with the lowest correlation with the dependent variable. The resulting ERS2 index was not correlated with subjective well-being, $r = 0.034$ ($p < 0.24$).

The analysis (Table 2) showed an interesting pattern, different for the ERS index correlated with the dependent variable (ERS1), and for the index uncorrelated with the DV (ERS2). The lack of correlation between ERS2 and subjective well-being turned out to be crucial to the results. ERS1 changed the model significantly already at STEP2, rendering conservatism – which was a key explanatory variable in the model – non-significant, and therefore changing the original conclusions of Napier & Jost. However, the ERS2 index, while by itself staying in the model as a close-to-significant ($p < 0.10$) predictor in all four steps of the analysis, did not cause any significant changes in the original model. It did not significantly influence any of the explanatory variables. This effect can be attributed to the specific topic of the analysis of Napier & Jost – conservatism. The concept of conservative behavior suggests that those experiencing it may also show a tendency for expressing their viewpoints in a more direct, well-defined manner – and as a result, they might more often provide extreme answers.

4.3. Study 2: Citizen-making: The role of national goals for socializing children

The second replication we conducted was a study of Bond & Lun (2014). Their paper reported on the profiling of 55 nations by two dimensions of the socialization goals for children extracted from the World Values Survey, *viz.*, self-directedness versus other-directedness, and civility versus practicality. We were mainly interested in the part about the mediating role of the two socialization goals in the relationship between a society's development, in terms of its Human Development Index, and its citizens' well-being, in terms of average satisfaction with life and average happiness. As the authors used data from World Survey Value wave 5, two new scales of 10 items were drawn, according to the methodology described in section 4.1 above, and two independent ERS indices were constructed.

At the country level, ERS index 2 was not significantly correlated with the dependent variables, but ERS index 1 was ($r_{satisfaction} = –0.4$, $p < 0.01$; $r_{happiness} = –0.29$, $p < 0.05$).

Predictor	**Step 1**	**S1+ ERS1**	**S1+ ERS2**	**Step 2**	**S2+ ERS1**	**S2+ ERS2**	**S3**	**S3+ ERS1**	**S3 ERS2**	**S4**	**S4+ ERS1**	**S4+ ERS2**
Constant	0.74***	0.70***	0.73***	0.65***	0.61***	0.64***	0.64***	0.6***	0.63***	0.59***	0.57***	0.59***
Conservatism	**0.07****	**0.05***	**0.072****	**0.06***	**0.04**	**0.06***	**0.05***	**0.03**	**0.05+**	**0.04+**	**0.03**	**0.04**
Income				0.05***	0.06***	0.05***	0.05***	0.05***	0.05**	0.05***	0.05***	0.05***
Education				0.02	0.03	0.02	0.02	0.02+	0.02	0.01	0.02	0.01
Sex				0.004	0.002	0.006	0.001	0	0.002	0.002	0.001	0.003
Age				0.007	–0.001	0.005	0.004	–0.003	0.003	0.004	–0.002	0.003
Age squared				2.9E-9*	3.0E-9*	2.8E-9*	2.8E-9*	2.5E-9*	2.7E-9*	2.5E-9*	2.3E-9+	2.4E-9+
Marital status				0.08***	0.07***	0.08	0.08***	0.07***	0.08***	0.07***	0.07***	0.07***
Employment				–0.05*	–0.05*	–0.05*	–0.05*	–0.05*	–0.05*	–0.05*	–0.04*	–0.05*
Church attend.							0.04*	0.02	0.03*	0.3*	0.02	0.03*
Meritocratic beliefs										0.08***	0.06***	0.07***
ERS		0.13***	0.03		0.15***	0.05+		0.14***	0.05+		0.13***	0.05+
Adj. R-squared	0.008	0.023		0.079	0.098	0.081	0.082	0.099	0.092	0.094	0.107	0.096

+ p < .10; * p < .05; ** p < .01; *** p < .001

Tab. 2. Results of regression analysis for study "Why conservatives are happier than liberals", as reported by Napier & Jost (2008), and with ERS added to each stage of the analysis

DOI 10.7172/1644-9584.60.4

We faced minor reproducibility problems when repeating the analysis – it was not clear whether data from Hungary was included in the analysis. Following the authors, running PROCESS, we analyzed mediation models with 5000 bootstrapping replications and the bias-corrected (BC) confidence intervals using regression for testing indirect effects. Controlling for ERS did not alter the results, neither for "satisfaction" nor for "happiness", on neither of the two dimensions: self-directedness versus other-directedness, and civility versus practicality (Table 3).

Model	Indirect effect	Estimate; BC 95% CI (lower, upper)		
		No ERS control	ERS1	ERS2 (not correlated with DV)
Life satisfaction	Self-directedness vs other-directedness	**1.33** (0.30;2.80)	**0.91** (0.18; 2.08)	**1.31** (0.40; 2.75)
	Civility vs practicality	**0.63** (0.09; 1.59)	**0.78** (0.11; 1.95)	**0.80** (0.15; 1.89)
	Total	**1.96** (0.75; 3.53)	**1.70** (0.65; 3.23)	**2.10** (0.94; 3.60)
Happiness	Self-directedness vs other-directedness	**0.49** (0.11; 0.96)	**0.35** (0.56; 0.73)	**0.45** (0.12; 0.89)
	Civility vs practicality	**0.21** (0.03; 0.50)	**0.26** (0.41; 0.63)	**0.26** (0.55; 0.58)
	Total	**0.70** (0.26; 1.18)	**0.60** (0.22; 1.07)	**0.72** (0.28; 1.15)

Tab. 3. Summary of indirect effects of HDI. All effects significant at the 0.05 level

4.4. Study 3: Examining the relation of religion and spirituality to subjective well-being across national cultures

The third replication were Lun & Bond's analyses from 2013. The authors tested a relation between different measures of spirituality and psychological well-being using data from wave 5 WVS. They describe inconclusive results on the subject in previous literature and apply several different measures of spirituality to discern which of the approaches to spirituality in particular are related to psychological well-being. Additional variables of HDI (Human Development Index), SHI (Social Hostilities Index) and support for religious socialization were added in the analyses, broadening the inquiries to include country level as well as individual level. We attempted to replicate Lun & Bond's (2013) hierarchical linear model analysis of all the aforementioned variables.

The article described analyses on the same wave of Word Value Survey, and the same dependent variables (happiness and life satisfaction), as the one replicated in the previous section. We decided to use the same ERS measures; in this case, however, analyses were conducted at the individual level and both ERS indices correlated significantly with the dependent variables. We repeated the procedure of random items draw and attempted

to create an index (ERS2) not correlated with the dependent variables (Table 4), but did not achieve the desired effect. With a sample size over 80,000, even a tiny correlation coefficient will easily be found significant. Even purposeful selection of items would not produce a measure uncorrelated with happiness and life satisfaction on a sample of this size. We therefore proceeded with the analyses, cautious of the possibility that ERS effect might be confounded with other theoretical content involved in the analysis.

	Happiness	**Life satisfaction**
ERS index #1	.182**	.080**
ERS index #2	.073**	.048**

Tab. 4. Relationship between ERS indices and the dependent variables. All Pearson correlation coefficients significant at $p < 0.01$; $N = 83018$

We replicated most of the relationships the authors had found with happiness as the dependent variable. We confirmed the significance of the demographic variables in the model (age and social class) as well as significance of individual level variables: belief in religious authorities, value of religion and religious identity. Our replications were imperfect, however, as we observed relationships not present in the original analyses as well. Those were the significance of national level variables and of the social religious practice variable.

Including any of the ERS indices did not alter the analysis conclusions.

An attempt to replicate Lun & Bond's findings on the second measure of psychological well-being (satisfaction with life – Table 5) brought similar results. We replicated most but not all of the findings. At the individual level we observed the relationships between life satisfaction and: belief in religious authorities, value of God or gods, religious identity as well as a weaker relationship with the value of religion. These findings were congruent with the original analysis. At the national level we confirmed HDI as related to life satisfaction, but in our analyses SHI proved significant as well. Age and social class proved significant, while gender did not in both our and Lun & Bond's analyses. Among covariates, we failed to replicate the significance of the educational level.

Including a measure of ERS in our analyses altered the conclusion in the case of two of the fourteen observed variables for ERS index 2 less correlated with the DV. For the more correlated ERS index #1 the changes in relationships significance in the model included the educational level covariate as well.

 DOI 10.7172/1644-9584.60.4

Main effects	Original variables only		Including ERS index No. 1		Including ERS index No. 2	
	b	Std. Error	b	Std. Error	b	Std. Error
Covariates						
Age	–.017***	.003	–.020***	.003	–.01800***	.003
*Age*2	.000**	.000	.000***	.000	0.00009**	.000
Gender	–.043	.019	–.038	.019	–.03100	.019
Education level	–.007	.005	–.012*	.005	–.01100	.005
Social class	–.486***	.010	–.485***	.010	–.48500***	.010
ERS	—	—	1.573***	.047	.94600	.051
Individual level						
Belief in religious authorities	–.204***	.027	–.231***	.027	–.20600***	.027
Value of God or the gods	.034***	.005	.022***	.005	.02900***	.005
Value of religion	.034**	.014	.031	.014	.02600	.014
Spiritual practice	–.055	.028	–.040	.027	–.04200	.028
Social religious practice (frequent)	.047	.024	.092***	.023	.07400***	.024
Religious identity	.439***	.026	.398***	.026	.42100	.026
National level						
HDI	2.927***	.076	3.259***	.075	3.08700***	.076
SHI	–.075***	.005	–.084***	.005	–.08500***	.005
Support for religious socialization	–.005***	.001	–.005***	.001	–.00500***	.001

* $p < .05$; ** $p < .01$; *** $p < 0.001$

Tab. 5. Satisfaction with life model

And as our ERS indices were correlated to both the independent and dependent variables, we could not avoid confounding the ERS effect with the meaning of the items, and with subject-level relationships between the measured constructs. This puts even more emphasis on selecting an ERS measure that is not related in any way to the DV at least and may pose yet another obstacle to including ERS in studies done on large data sets that include analyses at the individual level.

5. Theoretical model

The results of our analyses forced us to include a more theoretical approach to how ERS could influence the results of the analysis. An effect of the tendency to provide more extreme responses can be analyzed looking at:

- Correlations – can strong ERS result in increased or lowered linear relationships between variables?
- Means – can strong ERS result in higher / lower means on affected scales / compound indicators?

5.1. How ERS influences respondents' reactions

In order to further explore ERS in social studies and attempt to define the boundaries of this phenomenon, we have to move on from individual cases of study replication. In the following chapters we therefore propose a theoretical model that might explain some pertaining questions asked in this paper. Let us consider a simplistic case with two items forming a compound indicator A, and two different items forming an indicator B. Each item is supplied with a 5-point response scale (with answers: No, Rather no, Neither yes nor no, Rather yes, Yes). Although such response scales are usually coded as 1 through 5, to simplify the example we coded the categories as –2, –1, 0, 1, 2. This centering of the range does not influence the relationships between variables.

Items A1 and A2 are reverse-coded. High values on A1 have the same meaning as low values on A2. Indicator B is represented in the same manner. This is a common design technique to account for acquaint response bias, i.e. a tendency to agree with items regardless of their meaning.

A1	A2	B1	B2	
2	2	2	2	Yes
1	1	1	1	Rather yes
0	0	0	0	Neither yes or no
–1	–1	–1	–1	Rather no
–2	–2	–2	–2	No

Tab. 6. Four items for two compound indicators: A & B. The same color shows answer categories with a similar meaning within the indicator

People with a strong tendency to provide extreme responses will view this response scale in a slightly different way. For them, "rather yes" and "rather no" will not be important and will not be used often. Thus, for high-ERS respondents the scale is – internally, inside their minds – transformed into a structure presented in Table 7.

 DOI 10.7172/1644-9584.60.4

Formal representation of the scale				Response scale with unused categories removed			
A1	**A2**	**B1**	**B2**	**A1**	**A2**	**B1**	**B2**
2	2	2	2	2	2	2	2
1	1	1	1	0	0	0	0
0	0	0	0	–2	–2	–2	–2
–1	–1	–1	–1				
–2	–2	–2	–2				

Tab. 7. Four items for two compound indicators: A & B, for high-ERS respondents. The same color shows answer categories with a similar meaning within each indicator. White cells show response categories unused or rarely used by high-ERS respondents

Similarly, respondents who tend to avoid issuing strong, extreme responses, will use a response scale modified in such a way that "yes" and "no" categories are mostly unused, also reducing the originally 5-point scale to just three response categories – see Table 8.

Formal representation of the scale				Response scale with unused categories removed			
A1	**A2**	**B1**	**B2**	**A1**	**A2**	**B1**	**B2**
2	2	2	2	1	1	1	1
1	1	1	1	0	0	0	0
0	0	0	0	–1	–1	–1	–1
–1	–1	–1	–1				
–2	–2	–2	–2				

Tab. 8. Four items for two compound indicators: A & B, for low-ERS respondents. The same color shows answer categories with a similar meaning within each indicator. White cells show response categories unused or rarely used by low-ERS respondents

5.2. Effect on correlations

Subjects who prefer extreme responses are expected to provide them for a majority, if not for all, items. The order of responses stays unchanged, however, for both high- and low-ERS subjects. A linear correlation coefficient is used to detect whether high values on one variable (e.g. indicator A, in our example) systematically co-exist with high values on another variable (e.g. indicator B), resulting in a high positive correlation; or with low values, resulting in a high negative correlation.

Extreme response style does not influence the order of answers, and so high values on variable A will stay connected with high values on variable B, regardless of whether a person has a tendency for high or low extreme answers.

Interval level correlations

The effect of ERS on correlations should be analyzed in two scenarios: when the researcher uses ordinal-level statistic (e.g. Spearman's correlation coefficient) or an interval-level statistic (e.g. Pearson's correlation coefficient).

On a 5-point measurement scale, which is reduced by ERS effect to 3 points, there should be no difference, as this is a case equivalent to multiplying all responses by a constant (1 becomes 2, –1 becomes –2 in our case). The same happens with a 4- point response scale. Such a transformation will have no effect, even on an interval-level correlation coefficient[1].

We cannot discuss ERS for scales consisting of fewer than 4 response categories. (e.g. "yes", "rather yes", "rather no", "no"). Longer response scales with more than 5 categories will affect respondents' reactions on the edges of the scale, and in some cases we could expect a marginally different effect, but only for interval-level correlation measures.

Using simulated data, we have further confirmed that there are only some very specific scenarios where ERS in correlational analyses is potentially worth taking into account.

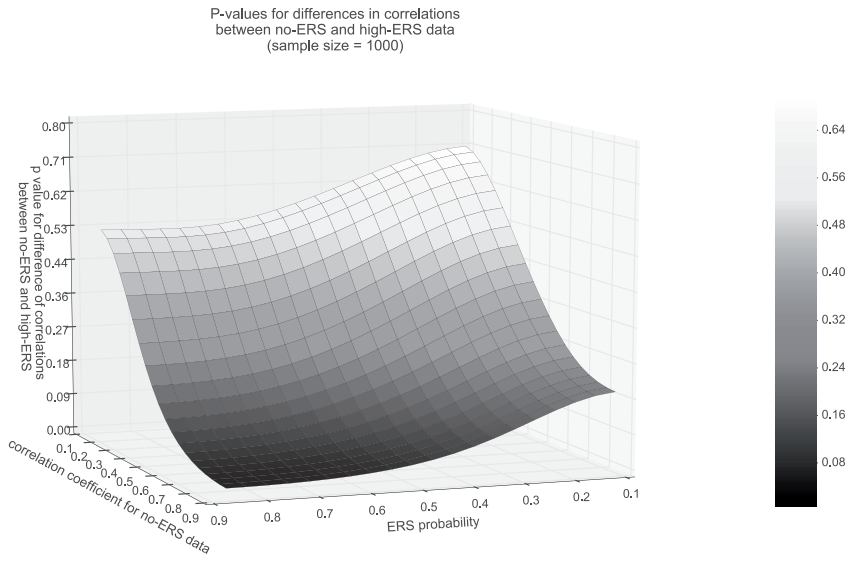

Fig. 1. Results of a simulation showing significance (p-values) for differences in correlation coefficients for samples of N = 1000 each, for various ERS probabilities in the population (horizontal axis 1), and for correlation coefficients for non-ERS data between 0.1 and 0.9 (horizontal axis 2)

 DOI 10.7172/1644-9584.60.4

As can be seen in Figure 1, showing simulation of two variables correlated in various degrees (between 0.1 and 0.9), for samples of size 1000, a near-significant difference between correlations for ERS-loaded data and non-ERS data can be potentially observed only where the correlation between the two variables is extremely high (approx. 0.7 or higher) and the probability of ERS influencing any data point is above approximately 0.5. In real world scenarios of social research, variables correlated at the level of 0.7 or higher are often part of a compound indicator. If such highly correlated variables are used in regression analyses as predictors, the high correlation will likely become source of multicollinearity, further reducing chances of seeing such data in analyses. Therefore, one additional area where ERS should be taken into consideration is its effect on reliability of compound indicators.

Further simulations, illustrated in Figure 2, show how significant changes in the value of correlation coefficient can be expected for different sample sizes and different probabilities of data points being influenced by ERS. It can be clearly seen that changes of potential statistical significance can be expected only for large sample sizes (approx. $N > 500$) and for high probability that any given data point will be influenced by the ERS effect (approximately > 0.7).

The simulations are illustrated not with raw values of r-change, but instead with p-values associated with an expected change – as raw r values would be incomparable, especially in the context of different sample sizes.

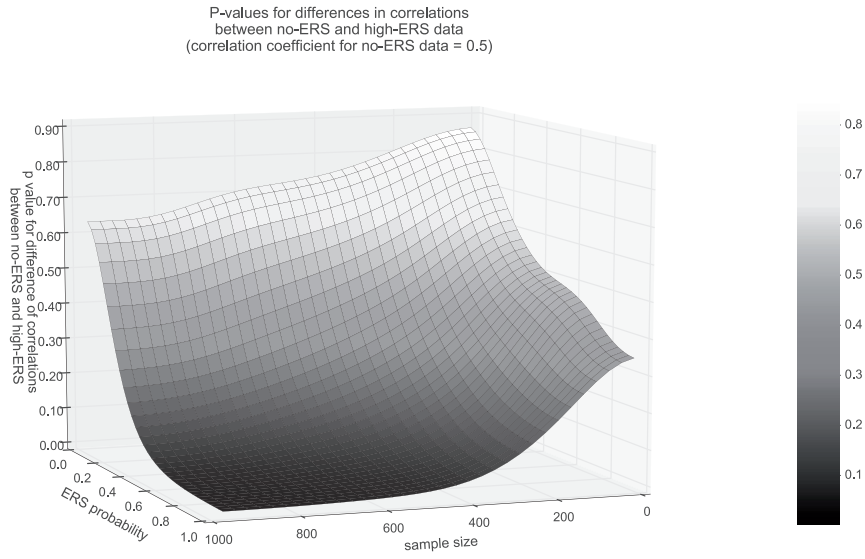

Figure 2. Results of a simulation showing significance (p-values) for differences in correlation coefficients, for samples sizes between 10 and 1000 (horizontal axis 1), for various probabilities of ERS in the population (horizontal axis 2), for samples drawn from a population with r = 0.5 (r value for data not influenced by ERS)

Ordinal level correlations

For any statistic operating on an ordinal measurement scale, exact numeric values are irrelevant, as long as the ordering stays intact. Shifting the response category towards more or less extreme, without affecting the ordering of answers, will – by definition – have no effect on the value of ordinal level statistics, e.g. *Spearman's r*.

5.3. Effect on between-group differences in averages

The lack of influence of ERS on correlations does not imply whether ERS may influence statistical analyses based on differences in average results between groups – a typical "experimental" design.

ERS will affect the results of mean comparison by increasing the between-group variance for respondents preferring to use extreme responses and by decreasing the between-group variance for respondents avoiding extreme responses. Furthermore, unless accounted for, it will increase within-group (unexplained) variance, as we can expect that in any group of respondents, a mixture of high- and low-ERS respondents will be found.

To illustrate the effect, we will use a simple example with four respondents, randomly assigned to two groups: group 1 (watched a movie about nature), and group 2 (watched a movie about technology). All respondents then answered a question "Would you be willing to give 1% of your income to protect the natural environment?"

In Table 9, results for the "high-ERS" respondents in the above example are presented. Such respondents, whether they have positive or negative views on a given subject, will avoid "rather yes" and "rather no" categories, and instead choose "yes" and "no" options. As a result, the between-group difference in means will be maximized ($AVG_{[grp1]} = 1$; $AVG_{[grp2]} = -1$). Comparatively, as can be seen in Table 10, the same scenario for low-ERS respondents will result in a lower average between-group difference ($AVG_{[grp1]} = 0.5$; $AVG_{[grp2]} = -0.5$).

Formal representation of the scale; categories preferred by high-ERS respondents highlighted

A1	A2	B1	B2	
2	2	2	2	Yes
1	1	1	1	Rather yes
0	0	0	0	Neither yes or no
–1	–1	–1	–1	Rather no
–2	–2	–2	–2	No

Response scale with unused categories removed

Person ID	GRP	A1
#1 "positive"	1	2
#2 "undecided"	1	0
#3 "negative"	2	–2
#4 "undecided"	2	0

Tab. 9. Example results for four high-ERS respondents in two groups. High-ERS respondents with a positive attitude to a given subject will prefer "Yes" category over "Rather yes", and a similar pattern can be found for respondents with a negative attitude to a given subject

 DOI 10.7172/1644-9584.60.4

Formal representation of the scale; categories preferred by low-ERS respondents highlighted

A1	A2	B1	B2	
2	2	2	2	Yes
1	1	1	1	Rather yes
0	0	0	0	Neither yes or no
–1	–1	–1	–1	Rather no
–2	–2	–2	–2	No

Response scale with unused categories removed

Person ID	GRP	A1
#1 "positive"	1	1
#2 "undecided"	1	0
#3 "negative"	2	–1
#4 "undecided"	2	0

Tab. 10. Example results for four low-ERS respondents in two groups. High-ERS respondents with a positive attitude to a given subject will prefer "Yes" category over "Rather yes", and a similar pattern can be found for respondents with a negative attitude to a given subject

In real data, however, we cannot expect either high- or low-ERS respondents to form the majority or all of the participants in any research, unless the factor that we think of is not really a personal preference level for extreme answers, but rather a cultural level, or subject level difference. For example, a characteristic trait of the Japanese language is its natural tendency to avoid strong, extreme statements. It has been shown by Wieczorkowska, Wierzbiński & Kuźmińska (2014) that Japanese respondents tend to avoid extreme responses more than Italians.

As a result, it can be expected that in a typical analysis of variance scenario, if respondents do not show a global bias towards either high or low ERS pattern of replying, the resulting influence of the ERS is expressed only with an increased within-group variance.

Person ID	GRP	A1
#1 "positive low ERS"	1	1
#2 "undecided low ERS"	1	0
#3 "negative low ERS"	2	–1
#4 "undecided low ERS"	2	0
#5 "positive high ERS"	1	2
#6 "undecided high ERS"	1	0
#7 "negative high ERS"	2	–2
#8 "undecided high ERS"	2	0

Tab. 11. Example results for a merged low- and high-ERS respondents result with an increased within-group variance. ($AVG_{[grp1]} = 0.75$; $AVG_{[grp2]} = -0.75$)

6. Conclusions

Extreme response style has gained a lot of attention from social scientists over the last 20 years. The group most interested in assessing its potential influence are researchers dealing with international and cross-cultural research, as language- and culture-induced differences are perceived as important determinants of the tendency to use extreme ends of response scales. The aim of our study was to check if ERS can inflate or deflate correlations between respondents' scores on various scales, and whether it can influence results of ANOVA designs. We argue that even for interval-level correlation measures, as used e.g. in the ordinary least squares regression analysis, the effect of ERS on real data is negligible, unless the ERS indicators are correlated with key variables in the model, especially with the dependent variable. To support our thesis, we replicated several analyses, previously published in social research or management literature, and further modified them by adding the ERS measure. We showed that the effect of an uncorrelated (with DV) ERS on such analyses, even if performed on huge datasets consisting of multi-national data of thousands of cases, remains negligible and thus does not need to be taken into consideration in any but very specific scenarios. For ERS indicators that show a significant correlation with a dependent variable in a model, we have shown that they may, but not in every case do, influence the analyses of the model significantly. It is expected that the effect results from confounding of the ERS effect with subject-level variables, at the individual or cultural/national level.

Our simulation results, where the only near-significant influence of ERS is present in very highly correlated items, suggest that one additional area where ERS should be taken into consideration is its effect on the reliability of compound indicators. Furthermore, our simulations have shown that ERS may potentially influence simple bivariate correlations only in very large samples, and for very strong levels of the ERS, where the probability of any data point being modified by ERS is at least 0.7.

However, ERS remains an important factor in cross-cultural research, in comparisons of different groups or in individual diagnostics of a person (e.g. in recruitment). It should be also noted that formal response style factors, like ERS and ARS, might pose varying levels of influence, depending on the mode of delivery of the survey.

Endnote

[1] Of course, this is the simplest scenario, not possible with real data, where we cannot assume that every respondent belongs either to a pure high-ERS or low-ERS group.

 DOI 10.7172/1644-9584.60.4

References

Austin, E.J., Deary, I.J. & Egan, V.(2006). Individual differences in response scale use: mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, 40, 1235–1245.

Baumgartner, H. & Steenkamp, J-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143–156.

Batchelor, J.H., Miao, C. & McDaniel, M.A. (2013). Extreme Response Style: A Meta-Analysis. Presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology.

De Beuckelaer, A., WeijtErs, B. &·Rutten, A. (2010). Using ad hoc measures for response styles: a cautionary note. *Qual Quant*, 44, 761–775, DOI 10.1007/s11135-009-9225-z.

Bond, M.H. & Lun, V.M. (2013). Examining the Relation of Religion and Spirituality to Subjective Well-Being Across National Cultures. *Psychology of Religion and Spirituality, 5*(4), 304–315.

Bond, M.H. & Lun, V.M. (2014). Citizen-making: The role of national goals for socializing children. *Social Science Research,* 44, 75–85.

Cronbach, L.J. (1946). A case study of the split-half reliability coefficient. *The Journal of Educational Psychology*, *37*(3), 473–480.

Cheung, G.W. & Rensvold, R.B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187–212.

Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior & Personality*, *15*(1), 137–152.

Crandall, J. (1973). Sex Differences in Extreme Response Style: Differences in Frequency of Use of Extreme Positive and Negative Ratings. *The Journal of Social Psychology*, 89, 281–293.

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

Engle, P.J. (2011). An Experimental Assessment of Extreme Response Style, Prepared for the American Politics Workshop, available online, 20.12.2014, https://moodle.ied.edu.hk/pluginfile.php/28242/mod_forum/attachment/20436/An_Experimental_Assessment_of_Extreme_Response_Style.pdf.

Frowin, F., Memmert, D. & Unkelbach, C. (2013). Calibration processes in a serial talent test. *Psychology of Sport and Exercise*, 4, 488–492.

Greenleaf, E.A. (1992). Measuring Extreme Response Style, *Public Opinion Quarterly*, 56, 328–51.

Grim, S. & Church, A. (1999). A Cross-Cultural Study of Response Biases in Personality Measures, *Journal of Research in Personality*, *33*(4), 415–441.

Harzing, A., Brown, M., Köster, K. & Zhao, S. (2011). Response style differences in cross-national research: Dispositional and situational determinants. *Management International Review, 52*(3), 341–363.

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.

Johnson, T.R., Kulesa, P., Cho, Y.I. & Shavitt, S. (2005). The relation between culture and response styles: evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36, 264–277.

Kieruj, N.D. & Moors, G.(2013). Response style behavior: question format dependent or personal style? *Qual Quant*, 47, 193–211, DOI 10.1007/s11135-011-9511-4.

Lau, M.Y. (2008). *Extreme response style: An empirical investigation of the effects of scale response format and fatigue*, doctoral thesis, University of Notre Dame. Retrieved from https://books.google.pl/books?hl=pl&id=f8qePyfQbhwC&q= (10.09.2015).

Marshall, R. & Lee, C. (1998). A Cross-Cultural, Between-Gender Study of Extreme Response Style. *European Advances in Consumer Research*, 3, 90–95. Retrieved from http://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=11158.

Meisenberg, G. & Wiliams, A.(2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44, 1539–1550.

Napier, J.L. & Jost, J.T. (2008). Why are conservatives happier than liberals? *Psychological Science, 19*(6), 565–572.

Naemi, B.D., Beal, D.J. & Payne, S.C. (2009). Personality predictors of extreme response style. *Journal of Personality,* 77, 261–286.

Nisbett, R.E. (2009). *Geografia myślenia*. Sopot: SAW Smak Słowa.

PetErson, R.A., Rhi-Perez, P. & Albaum, G. (2014). A cross-national comparison of extreme response style measures. *International Journal of Market Research, 56*(1), 89–110.

Roster, C.A., Albaum, G. & Rogers, R.D. (2006). Can cross-national/cultural studies presume etic equivalency in respondents' use of extreme categories of Likert rating scales? *The Market Research Society*, *48*(6),741–759.

Warnecke, R.B., Johnson, T.P., Chávez, N., Sudman, S., O'Rourke, D.P., Lacey, L. & Horm, J. (1997). Improving question wording in surveys of culturally diverse populations. *Ann. Epidemiol.*, 7, 334–342.

Wieczorkowska, G. (1993). Pułapki statystyczne. In: Z. Smoleńska (ed.), *Badania nad rozwojem w okresie dorastania* (pp. 211–234). Warszawa: PAN.

[WWK] Wieczorkowska-Wierzbińska, G., Wierzbiński, J. & Kuźmińska, A. (2014). Porównywalność danych sondażowych zebranych w różnych krajach. *Psychologia Społeczna, 2*(29), 128–143.

 DOI 10.7172/1644-9584.60.4