

Article

Cultural Differences in Careless Responding

Journal of Cross-Cultural Psychology
2019, Vol. 50(3) 336–357
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0022022119827379
journals.sagepub.com/home/jcc



Ina Grau¹, Christine Ebbeler¹, and Rainer Banse¹

Abstract

Careless responding (CR) in surveys has been identified as a serious threat for the validity of survey data. It occurs when participants respond without following the instructions or reading the content of items. The purpose of this article is to assess and explain the amount of CR at the individual and country level. In our study, individual CR (measured by indices using a short Big 5 personality questionnaire) correlated with education level and personality traits. In addition, using a sample of 8,320 participants from 34 countries, CR at the country level was investigated and it was strongly correlated with the cultural dimensions: human development, individualism, gender inequality, and power distance. CR can be seen as a powerful predictor of differences between countries. Finally, a comparison between CR and response styles (extreme and midpoint answers, acquiescence, and socially desirable responses) was conducted. CR and response styles showed some overlap, but loaded on different factors. CR is compatible with extreme answers, midpoint answers, and acquiescence, but different from social desirability. Social desirability is the only response style which requires careful reading of items.

Keywords

careless responding, response sets, response styles, cross cultural comparison, Big 5

Keywords

careless responding, response sets, response styles, cross-cultural comparison, Big 5

Researchers rely on research participants who truthfully respond to numerous survey items. We normally assume that respondents may present themselves in a socially desirable way but that they respond to the questions asked. However, sometimes some people do not, a phenomenon called careless responding (CR). CR occurs when respondents fail to read instructions or items carefully (Bowling et al., 2016), which leads to monotonous or inconsistent answer patterns. CR is normally treated as an individual-level variable caused by a minority of a sample. In the present study, CR was investigated also at the country level in relation to cultural dimensions.

CR is a content-free response set that is normally conceived of as a result of a lack of motivation to read and to answer to the survey questions correctly (Bowling et al., 2016). CR can be measured using any questionnaire which includes at least one reverse-scored item per construct (e.g., a Big 5 questionnaire). Of course, the validity of the originally measured constructs can suffer from CR, but the amount of CR, which can be computed from the responses, constitutes a reliable and valid new construct on its own that is systematically related with personality traits. Bowling et al. (2016) have demonstrated that individual differences of CR were correlated

¹University of Bonn, Germany

Corresponding Author:

Ina Grau, Department of Psychology, University of Bonn, Kaiser-Karl-Ring 9, 53111 Bonn, Germany.
Email: igrau@uni-bonn.de

negatively with conscientiousness, agreeableness, extraversion, and emotional stability that were reported by an acquaintance of the participants.

In many studies, it is common practice to identify and drop 3% to 10% of the participants scoring high on CR with the goal of improving the reliability and validity of the results (e.g., Curran, 2016; Johnson, 2005; Meade & Craig, 2012). The impact of CR can be investigated by testing whether it moderates the correlation between variables of interest (Piedmont, McCrae, Riemann, & Angleitner, 2000). However, the usefulness of controlling for CR is controversial. Some authors call into question that dropping high CR respondents improves the results (Kurtz & Parrish, 2001; McGrath, Mitchell, Kim, & Hough, 2010; Piedmont et al., 2000), others found an impact of CR on data quality, correlations, factor structure, and statistical power (Huang, Curran, Keeney, Poposki, & DeShon, 2012; Maniaci & Rogge, 2014). However, in some of these studies, response sets like CR and response styles like social desirability are studied simultaneously, and the impact of CR alone remains unclear. In addition, most studies were conducted with student samples. In the present study, we examine CR using a nonstudent sample in 34 countries. This approach allows for studying CR simultaneously at the individual and country level.

The first purpose of this study is to introduce a composite index of CR combining five facets and to illustrate CR's correlations with variables such as education level and the Big 5 personality traits across a large number of countries. Second, we investigate how cross-cultural differences in CR are related to well-established cultural dimensions such as human development, individualism, gender equality, and equality between social groups. Third, we compare the CR facets with well-established facets of response styles and demonstrate overlap and differences between CR and these response styles.

Facets of CR

A large number of indices measuring CR have been described in previous research (e.g., Curran, 2016; Desimone, Harms, & DeSimone, 2014; Johnson, 2005; Meade & Craig, 2012). The differ-

Facets of CR

A large number of indices measuring CR have been described in previous research (e.g., Curran, 2016; Desimone, Harms, & DeSimone, 2014; Johnson, 2005; Meade & Craig, 2012). The different CR indices can be broadly classified into three types: direct, archival, and statistical (Desimone et al., 2014).

Direct screening methods require special items to assess CR, while archival and statistical methods can be derived post hoc from other measures. Self-reported indices of data quality, such as "I occasionally answered items without reading them" belong to the direct methods. The direct technique of instructed items or attention check items (Kung, Kwok, & Brown, 2018) requires a special reaction from the respondent, such as "Please leave this item blank." Bogus items such as "I was born on planet earth" have an unambiguous correct answer. A wrong response to this kind of questions is a clear signal of CR.

These types of methods for detecting CR are highly transparent for respondents. Respondents may or may not answer honestly to direct CR questions, but it is also possible that they feel that the researcher expects them to give invalid answers, which in itself can reduce the motivation to answer questions honestly. Another issue with direct CR screening methods is that the researcher has to plan a priori to assess CR. Archival and statistical techniques can be used post hoc to reanalyze already existing data sets.

The use of semantic synonyms is an archival technique to assess CR. Parallel items such as "I enjoy my job" and "I like my current occupation" (examples by Desimone et al., 2014) have to be answered identically or at least similarly. The opposite is true for semantic antonyms, which require different responses. It has to be noted that saying yes to two questions with opposed content is sometimes seen as an indicator of dialectical thinking (Peng & Nisbett, 1999) or a distinction between different situations (Hampson, 1998)—one can be generous in some situations (e.g., spending a lot of money for traveling) and thrifty in other situations (e.g., in buying clothes). However, if respondents give inconsistent answers, the items are mostly consistent regarding

evaluation (positive or negative). More importantly, in questionnaires such as the Big 5, it is not expected to find negative item-total-correlations for reverse-scored items. Thus, we consider opposed answers in the Big 5 traits as CR.

Longstrings are defined as monotonous answering behavior, or selecting the same option repeatedly, for example, always checking the "4" in a 5-point scale. The longest series of identical answers in a questionnaire can be used as an index. This technique is recommended if more than one construct and/or reverse-coded items are included in the questionnaire.

One of the statistical CR screening methods is the Mahalanobis (1936) distance, which is defined as the multivariate distance between the answer given by a specific person and the mean answer given by all other persons responding to the same questionnaire items. Participants showing a high Mahalanobis distance can be considered as multivariate outliers, as they show a different response pattern than the rest of the sample. A second statistical technique is the analysis of psychometric synonyms. This method is comparable to the use of semantic synonyms mentioned above. The difference is that semantic synonyms are obvious due to their content, while psychometric synonyms can be found by analyzing the correlations between items in a sample. Item pairs with a correlation of at least $r = .60$ can be seen as synonyms (Meade & Craig, 2012). If a questionnaire features three or more synonymous pairs of items, a high intraperson correlation across these pairs of items would indicate careful answering or a low score on CR. The same is true for a negative correlation of answers to psychometric antonyms. Finally, the person-total correlation presents a measure of the similarity of responses of each person with all other persons (Curran, 2016). In reliability analyses, the item-total correlation indicates how well one item can predict all other items. In contrast, the person-total correlation indicates how well a person can predict the answers of all other persons. Low person-total correlations are interpreted as an indicator of CR.

Other indices of CR which are not used in the current study are response time (in computer administered studies; cf. Meade & Craig, 2012), the technique of personal reliability (Johnson, 2005), interitem standard deviation (Curran, 2016), and the detection of polytomous Guttman errors (Curran, 2016).

Sometimes CR is found by chance. Piferi and Ickes (2003) presented a questionnaire with an

Sometimes, CR is found by chance. Piferi and Jobe (2003) presented a questionnaire with an extra answer sheet. Accidentally, the answer sheet had 10 more response scales than corresponding survey questions. However, out of 159 participants, 16 answered all of these 10 "ghost items," which showed that they had not read the items at all, at least at the bottom of the questionnaire. Another indication that the 10% who filled in the blank items did not pay much attention to the other items either was that the correlation between reverse-coded and not reverse-coded items (after recoding) was negative for the ghost-item responders, while it was positive for the majority.

Causes, Correlates, and Consequences of CR

Several causes of CR have been identified (Meade & Craig, 2012), namely, having a low level of interest in the survey, surveys being too long, a lack of social contact with the researcher, and environmental distractions. Regarding the intercorrelations between CR indices, Meade and Craig (2012) found three factors: The first factor contained psychometric antonyms and synonyms, the odd-even consistency, Mahalanobis distance, and bogus items. A second factor contained four self-reported items of effort in the study. A third factor contained two longstring indices. These results showed that CR is a multidimensional construct with a meaningful factor structure.

The indices of CR showed substantial correlations not only between each other (Meade & Craig, 2012) but also with personality variables (Bowling et al., 2016). Bowling et al. (2016) found that CR (measured as an overall index of individual reliability, psychometric synonyms, psychometric antonyms, and Mahalanobis distance after recoding individual reliability and psychometric synonyms) was negatively correlated with acquaintance-reported conscientiousness, agreeableness, extraversion, and emotional stability. CR showed good stability over 13 months

and predicted motivational and achievement factors in academic studies (e.g., class absence and grades). The authors concluded that CR is "in part a manifestation of enduring individual differences" (p. 218).

*CR can distort results in any direction, it can inflate or deflate correlations between constructs (Kam & Meyer, 2015). Careless respondents produce not only unexpected correlations but also unexpected factor structures. Kam and Meyer (2015) showed that items measuring job satisfaction and job dissatisfaction produced a one-factor solution, but including data sets with high CR and a high level of acquiescence led to a two-factor solution. Woods (2006) showed in a simulation study including positively and negatively worded questions that the expected one-factor-solution fitted worse if 10% or more questionnaires were completed randomly. Instead, a second (method-)factor occurred. Furthermore, CR can change mean values of constructs. If the mean value of a sample is above the scale midpoint, careless respondents will reduce the mean value; if the mean value is below the scale midpoint, CR will increase the mean value (Huang, Liu, & Bowling, 2015).

In sum, mean values, reliability, factor structure, and correlations between constructs can be influenced by CR. Obviously this problem should not be ignored. One option is to drop participants showing extreme levels of CR from the analysis or to try to avoid CR, for example, by offering an incentive for careful responding (Bowling et al., 2016). In the present study, however, we try to explain the variance of CR in a more basic research approach.

We assume that CR is a reliable person variable that is related to personality traits. Conscientiousness is expected to correlate negatively with CR due to the conceptual overlap between the constructs conscientiousness and carefulness. Agreeableness reflects a tendency to be altruistic and to help the researcher. Therefore, it is also expected to correlate negatively with CR. In the study by Bowling et al. (2016), openness was unrelated to CR, and extraversion and emotional stability were negatively correlated with CR. We therefore expect to find the same pattern. Furthermore, a long school or university career might increase familiarity with questionnaires, which should result in less CR. We expect education level to be negatively related to CR. Sex and age are not expected to correlate with CR, but these variables are controlled for in the analysis.

CR in Different Countries

CR in Different Countries

It can be assumed that there are systematic differences in CR between countries as it is the case for response styles, for example, socially desirable, acquiescent, extreme, and midpoint responding. He, Bartram, Inceoglu, and Van de Vijver (2014) and He, van de Vijver, Dominguez Espinosa, and Mui (2014) found that the use of response styles was negatively related to the human development index (HDI) and individualism (IDV).

CR, however, was not examined interculturally in this way. In this study, we expect that differences in CR between countries can also be explained by cultural dimensions. Like He, Van de Vijver, et al. (2014), we included the HDI and IDV (as provided by Hofstede, 2001). The HDI is a composite measure of indicators along the three dimensions life expectancy, educational attainment, and per capita income (United Nations Development Programme, 2013). It is expected that CR is negatively correlated with the HDI. A high HDI reflects a higher level of education, which can make it easier for people to handle questionnaires. IDV (Hofstede, 2001) was also examined by He, Van de Vijver, et al. (2014) and showed the same effects as the HDI. IDV refers to the amount of interdependency between members of a specific society. It is expected that in individualistic countries the amount of CR is lower than in collectivistic countries. People in individualistic countries are possibly more used to being concerned with personal attitudes, feelings or evaluations, and might also be more prone to provide valid answers when asked about their personality and attitudes in a questionnaire.

The gender inequality index (GII) and the power distance index (PDI) were also included as explaining variables. Both are negatively related to the HDI and IDV. The GII combines the three

dimensions of reproductive health (maternal mortality ratio and adolescent birthrate), empowerment (share of parliament seats held by women, amount of women with at least secondary education), and labor market participation (female participation in the labor force), with higher values indicating higher inequality between men and women (United Nations Development Programme, 2013). The PDI (Hofstede, 2001) indicates to what extent power in a society is equally distributed among all society members, as opposed to some members of the society standing at the top of a hierarchical system and others at the bottom. There is no obvious theoretical reason to expect that GII and PDI are correlated with CR. However, because both cultural dimensions are related to the HDI, we expect them to be positively related to CR. The correlations between cultural variables raise the question of whether it is adequate to analyze them one by one.

Previous research has found high correlations between the cultural factors GII, HDI, IDV, and PDI (e.g., Archer, 2006; Ebbeler, Grau, & Banse, 2017; Fischer & Manstead, 2000). For this reason, Ebbeler et al. (2017) introduced a new superordinate factor that combines all four dimensions. Based on sociological terminology, the factor was called "emancipation," as an emancipated society can be defined as one in which "previously marginalized individuals and groups are free to fully participate across all spheres of society; one in which freedom rather than domination is evident in social and institutional practices" (cf. Dillon, 2014, p. 527). Alternative labels could be "level of democracy" or "open society." To solve the problem of conceptual overlap and statistical multicollinearity, the present research will also use the combination of all four cultural indices into this superordinate factor of emancipation. Emancipation is expected to correlate negatively with CR because it may reflect a higher education level, a higher familiarity with questionnaires and psychological research, and more opportunities to express one's opinion.

Comparing CR and Response Styles

Research on response styles in questionnaires has a long tradition. Response styles are defined as reactions of respondents to something other than the content of an item (Paulhus, 1991). It has

Comparing CR and Response Styles

Research on response styles in questionnaires has a long tradition. Response styles are defined as reactions of respondents to something other than the content of an item (Paulhus, 1991). It has been distinguished between social desirability (Eysenck & Eysenck, 1963), preferring extreme (or midpoint) answers (Shulman, 1973), and acquiescence, the tendency to rather say "yes" than "no," independently of the content of the item (Cronbach, 1946). Response styles are not seen as a random answering process but rather as personality traits correlated with age, sex, or social status (Ahammer, 1971; Couch & Keniston, 1960). As opposed to CR, which is seen as a result of lack of motivation, response styles are considered as a motivation to present oneself in a certain way. However, giving extreme, midpoint or yes-answers can also be done without reading the items. Therefore, we expect positive correlations between CR and the tendency to give extreme or midpoint answers and acquiescence. In contrast, socially desirable answering requires carefully reading the item content. Social desirability is therefore expected to correlate negatively with CR. He, Van de Vijver, et al. (2014) found that a composite Response Style Index at the country level correlated negatively with socioeconomic development and individualism. According to Smith (2004), the acquiescent response style is also more common in countries with high levels of power distance. Therefore, we expect the same correlations for response styles as for CR except for social desirability.

Hypotheses

Hypothesis 1: At the individual level, CR is expected to correlate negatively with education level, extraversion, agreeableness, and conscientiousness and positively with neuroticism. This hypothesis replicates and extends the results of Bowling et al. (2016) within a nonstudent sample from 34 countries. In addition, CR is expected to correlate negatively with the countries' emancipation level.