

Hypothesis 2: At the country level, CR is expected to correlate positively with PDI and GII and negatively with HDI, IDV, and the Emancipation Index. The same correlations are expected for response styles except for social desirability. The second part of the hypothesis refers to a replication and extension of the findings by He, Bartram, et al. (2014) and He, Van de Vijver, et al. (2014) by using more cultural dimensions and by distinguishing between social desirability and other response styles.

Hypothesis 3: The CR Index is expected to correlate positively with extreme answers, mid-point answers, and acquiescence and negatively with social desirability. The degree of overlap between CR and response styles will be explored with a factor analysis at the individual and at the country level.

Method

All participants who took part in the study were informed that participation was voluntary and all answers were anonymous. Participants were free to stop filling in the questionnaire at any point. Research ethics were explained in the funding application and reviewed and approved by the German Research Association. Therefore, no additional research ethics committee was involved.

Sample

The data presented here are a part of a cross-cultural study about predictors of marital satisfaction and functioning in 34 countries. Participants were married, but to avoid problems of anonymity only one spouse from each couple was recruited for the present study. After excluding 155 individuals who had missing values in any of the studied CR variables (at the scale level), the sample size was $N = 8,320$ from 34 countries (see Table 1), 54.23% participants were female. The mean age was 42.56 years ($SD = 12.71$ years, range = 18-91 years), and the mean marriage duration was 15.77 years ($SD = 12.42$, range = 1 month-70 years). Detailed information on the samples for each country is presented in Table 1.

The sample showed a tendency toward an overrepresentation of higher education in all participating countries. Especially in countries with relatively low literacy rates like Bangladesh or

Sample

The data presented here are a part of a cross-cultural study about predictors of marital satisfaction and functioning in 34 countries. Participants were married, but to avoid problems of anonymity only one spouse from each couple was recruited for the present study. After excluding 155 individuals who had missing values in any of the studied CR variables (at the scale level), the sample size was $N = 8,320$ from 34 countries (see Table 1), 54.23% participants were female. The mean age was 42.56 years ($SD = 12.71$ years, range = 18-91 years), and the mean marriage duration was 15.77 years ($SD = 12.42$, range = 1 month-70 years). Detailed information on the samples for each country is presented in Table 1.

The sample showed a tendency toward an overrepresentation of higher education in all participating countries. Especially in countries with relatively low literacy rates like Bangladesh or India, recruiting a more representative sample was not possible because it was mandatory for participants to complete the questionnaire on their own so that reliable data on the personal and potentially sensitive issues could be collected.

Procedure

Participants were recruited using different methods in different countries: by students of the University of Bonn who contacted participants during a stay abroad (i.e., while studying at a foreign university or visiting their country of origin), by market research institutes, or by cooperation partners from local psychology departments. Although market research institutes and cooperation partners aimed to recruit 150 married men and 150 married women (who were not married to each other), students recruited fewer participants (approximately 80-100). Fifteen of the countries in which data were collected were the countries of origin of foreign students or countries the students traveled to. The criteria for selecting the other 19 countries were extremely high or extremely low values on the United Nations' dimensions HDI and GII, as well as Hofstede's (2001) dimensions IDV and PDI. The recruitment strategies depended on the cultural norms of each country. For example, door-to-door recruitment was appropriate in some countries while approaching participants in central locations such as shopping malls or waiting rooms

Table 1. Descriptive Statistics.

Country	Language	Recruit- ment method	N	% female	Age	Education	Cronbach's α CR	CR index	Emancipation index
Ecuador	Spanish	1	298	53.0	37.3	4.5	.35	0.54	-0.89
Azerbaijan	Azerbaijani	1	79	51.9	41.3	5.6	.02	0.51	-0.49
Guatemala	Spanish	1	305	48.5	41.7	3.9	.46	0.38	-1.48
Pakistan	Urdu	1	278	49.3	36.2	5.0	.37	0.29	-1.18
Brazil	Portuguese	1	112	33.9	38.0	4.5	.29	0.22	-0.46
India	Hindi	2	400	50.0	39.7	3.3	.48	0.22	-0.99
Malaysia	Malay	1	446	53.4	38.8	3.7	.54	0.20	-0.58
Saudi Arabia	Arabic	1	311	58.8	31.1	2.6	.38	0.20	-0.79
Morocco	Arabic	1	100	51.0	40.1	3.5	.42	0.16	-0.64
Bulgaria	Bulgarian	1	217	71.9	43.9	4.2	.41	0.14	-0.14
Turkey	Turkish	1	175	53.7	41.3	4.3	.45	0.14	-0.35
Portugal	Portuguese	2	282	51.4	47.6	4.4	.12	0.11	0.06
Philippines	English	1	89	71.9	40.9	4.1	.41	0.05	-0.87
Georgia	Georgian	1	75	74.7	35.6	4.5	.33	0.05	-0.63
Greece	Greek	2	270	48.5	46.1	4.1	.38	0.02	0.27
Spain	Spanish	1	260	56.5	46.3	4.2	.37	0.01	0.58
Bangladesh	Bangla	2	300	50.0	38.3	5.7	.48	0.00	-1.28
Singapore	English	2	294	50.3	45.9	4.8	.37	-0.04	0.09
Romania	Romanian	1	364	54.9	41.7	5.1	.48	-0.06	-0.47
Japan	Japanese	3	388	53.1	49.0	4.0	.40	-0.07	0.56
Jamaica	English	2	309	60.2	44.0	4.5	.58	-0.07	-0.24
Slovakia	Slovak	3	298	48.0	48.0	4.1	.39	-0.09	-0.03
Russia	Russian	1	152	40.1	38.0	4.8	.44	-0.13	-0.37
Israel	Hebrew	1	244	51.6	35.8	4.9	.47	-0.16	1.00
Austria	German	1	211	51.7	49.4	4.2	.49	-0.20	1.07
Sweden	Swedish	1	354	55.6	44.0	4.4	.46	-0.22	1.17

Germany	German	1	361	68.4	41.1	4.3	.41	-0.24	1.07
Norway	Norwegian	1	165	63.6	45.7	4.3	.42	-0.27	1.21
Netherlands	Dutch	3	343	53.6	45.9	4.5	.50	-0.28	1.23
Latvia	Russian	1	103	56.3	38.3	4.5	.37	-0.29	0.62
Australia	English	3	252	48.8	53.8	4.7	.49	-0.30	1.30
The United States	English	1	309	51.5	46.7	4.4	.55	-0.32	1.08
Moldova	Russian	1	93	63.4	35.2	3.9	.44	-0.32	-0.61
New Zealand	English	1	83	69.9	46.3	3.8	.50	-0.37	1.20
Total/Average			8,320	54.2	42.6	4.3	.50	0.00	0.00

Note. Table is sorted by the CR Index. Higher values indicate more CR. Recruitment Method 1 = public space, 2 = door to door, 3 = panel. CR = careless responding.

Questionnaires were completed by the participants themselves, in their homes or at quiet places in public locations. Participants were assured that all information would be treated as anonymous. Completed questionnaires were returned in neutral, sealed envelopes, and submitted by mail, by putting them in special boxes, or handing them directly to a researcher. Incentives were given in some countries according to local social norms and recruitment method.

Measures

The test battery was translated into the local language using the translation-back-translation method (Brislin, 1980).¹ Translators were always native speakers of the language into which the

questionnaire was translated. Questionnaires were always adapted to local language forms (e.g., the Spanish version of the questionnaire for Guatemala) and approved by the respective cooperation partner in the country. The questionnaire battery included several scales regarding marital functioning, attitudes and behavior toward one's own marriage, and standard demographic characteristics. Demographic information included age, sex, number of children, and education level. Education level was assessed via a 6-point scale with answer possibilities adapted to the respective country (e.g., no school certificate, primary school, secondary school, upper secondary school level, apprenticeship, university).

Big 5 personality traits. To assess personality traits, a shortened version of the questionnaire by Benet-Martinez and John (1998) was used.² Each scale contained four items. The item selection was discussed with cooperation partners from India and Saudi Arabia to ensure that the wording was appropriate in these non-Western countries. Two items in the extraversion scale were reverse-coded, and one item in each of the remaining four scales (conscientiousness, agreeableness, neuroticism, and openness) was reverse-coded relative to the other items in each scale. Reliability analyses showed that the item-total correlations of the reverse-coded items (except extraversion) were very low (due to CR). So, we dropped the reverse-coded items from four of the five scales; however, these items were used to compute the CR measures. For the resulting Big 5 scales, Cronbach's α ranged from .66 (extraversion) to .79 (openness).³

CR. Measures of CR were taken from the Big 5 personality traits (as in other studies, for example, Meade & Craig, 2012). As some of the CR measures are redundant (semantic synonyms, psychometric synonyms) and some measures require special scale characteristics which were lacking in our study (e.g., Guttman errors), we used five indicators as a measure of CR: longstring, Mahalanobis distance, psychometric synonyms, psychometric antonyms, and person-total correlation. Longstring was coded as the longest row of identical answers in the 20-item Big 5 questionnaire. The Mahalanobis distance was calculated as the multivariate distance between a person's score

CR. Measures of CR were taken from the Big 5 personality traits (as in other studies, for example, Meade & Craig, 2012). As some of the CR measures are redundant (semantic synonyms, psychometric synonyms) and some measures require special scale characteristics which were lacking in our study (e.g., Guttman errors), we used five indicators as a measure of CR: longstring, Mahalanobis distance, psychometric synonyms, psychometric antonyms, and person-total correlation. Longstring was coded as the longest row of identical answers in the 20-item Big 5 questionnaire. The Mahalanobis distance was calculated as the multivariate distance between a person's score and the mean value of all other persons' scores for the 20 Big 5 items. For computing psychometric synonyms, four item pairs that approximately met the criterion of $r = .60$ were chosen (correlations were .57 for "outgoing-talkative," .59 for "do a thorough job-do things efficiently," .54 for "worry a lot-get nervous easily," and .69 for "inventive-original"). For each participant, an intraperson correlation was computed across these four pairs of items, with a high value indicating that the participant gave similar answers to synonyms and at the same time gave different answers to different constructs. The psychometric antonyms method usually requires item pairs with high negative correlations stronger than $r = -.60$, which were not found in the data. However, as the purpose of this analysis was not to define a cut-off point for dropping respondents from the analysis, smaller negative correlations were accepted for computing psychometric antonyms. To find stable negative correlations, which did not appear between item pairs, some items were aggregated: The mean of the two reverse-coded extraversion items (shy, quiet) correlated with the mean of the two nonreverse-coded extraversion items (outgoing, talkative) with $r = -.29$ over the whole sample. The reverse-coded neuroticism item (relaxed) correlated with the mean of the three nonreverse-coded items with $r = -.20$, and the equivalent correlation for conscientiousness (lazy) was $r = -.19$. From these three pairs, the intraperson correlation was computed, with a high negative correlation indicating careful responding, and a correlation near zero or even a positive correlation showing that reverse-coded items were overlooked. The low absolute value of these correlations shows that the performance in detecting reverse-coded items was relatively poor across the whole sample. Finally, the person-total correlation across all 20 Big 5 items was investigated as a measure of how well each person predicted the answer pattern of all other persons. For this calculation, the data set was transformed from a person-by-item matrix into an item-by-person matrix. The person-total correlation was computed in a reliability analysis.

The facets psychometric synonyms, psychometric antonyms, and longstring can be computed for each person individually. The two indices Mahalanobis distance and person-total correlation require a comparison of a person with the sample. As we intended to study cultural differences, we compared the participants with all other participants from all countries instead of comparing them with the persons from the same country. However, the differences between the methods are small. The correlation between the Mahalanobis distance of the participants compared with all other participants and the Mahalanobis distance of participants compared with only the persons from the same country was $r = .82$. The correlation of the person-total correlation based on all participants and the person-total correlation based only on other participants from the same country was $r = .88$.

An individual composite index of CR was computed as the mean value of the five standardized indicators for CR after recoding psychometric synonyms and person-total correlation (in these two variables high values indicate low CR). The CR index ranged from -1.02 to 2.43 with higher values indicating more CR. The reliability of this composite index was Cronbach's $\alpha = .50$ over all participants in all countries.⁴

Response styles. The response styles were also computed from the Big 5 questionnaire. To assess extreme answers, the frequency of choosing the endpoints (e.g., 1 and 5 in a 5-point scale) was computed. The tendency to give midpoint answers is the frequency of choosing the middle category (e.g., 3 in a 5-point scale). Acquiescence was computed as the mean value of all 20 Big 5 items without recoding reverse worded items. He, Van de Vijver, et al. (2014) used the frequency of choosing the second highest alternative (for the purpose of avoiding a confusion with extreme answers) instead. The authors found factor loadings near zero for acquiescence on a response style factor and dropped acquiescence from their response style factor in later studies (He & Van de Vijver, 2015). Therefore, we chose the alternative computation as a continuous measure for

Response styles. The response styles were also computed from the Big 5 questionnaire. To assess extreme answers, the frequency of choosing the endpoints (e.g., 1 and 5 in a 5-point scale) was computed. The tendency to give midpoint answers is the frequency of choosing the middle category (e.g., 3 in a 5-point scale). Acquiescence was computed as the mean value of all 20 Big 5 items without recoding reverse worded items. He, Van de Vijver, et al. (2014) used the frequency of choosing the second highest alternative (for the purpose of avoiding a confusion with extreme answers) instead. The authors found factor loadings near zero for acquiescence on a response style factor and dropped acquiescence from their response style factor in later studies (He & Van de Vijver, 2015). Therefore, we chose the alternative computation as a continuous measure for agreement. For the assessment of social desirability, we did not implement an extra social desirability scale, as social desirability is also included in personality questionnaires. According to Paulhus (1981), the first unrotated factor of a multidimensional questionnaire measures social desirability, because the positive or negative valence is what all items of the Big 5 have in common. A factor analysis using the 20 Big 5 items was conducted. For each person, the factor value of the first unrotated factor was used as an index for social desirability. All items of the Big 5 questionnaire loaded positively on this factor except for the neuroticism items which are socially undesirable. He, Van de Vijver, et al. (2014) introduced a common response style factor with a positive loading of social desirability and extreme answers, a negative loading of midpoint answers, and a low loading of acquiescence. We changed the computation of acquiescence as described before and found a positive loading of acquiescence (0.56) instead. The loading of extreme answers was 0.80, midpoint answers -0.78 , and social desirability 0.82. The eigenvalue of the first factor was 2.23, explaining 55.78% of variance. The response style index was computed as the mean value of these four facets after recoding the tendency to give midpoint answers (Cronbach's $\alpha = .60$). To ensure that the four response styles are included in the index with the same weight, they were previously z transformed.

Variables at country level. PDI and IDV for the countries studied were taken from Hofstede (2001).⁵ As a measure of gender equality we used the GII, and as a measure of economic development the HDI as provided by the Human Development Report (United Nations Development Programme, 2012). Due to the very high correlations between these four cultural dimensions (see

Variables at country level. PDI and IDV for the countries studied were taken from Hofstede (2001).⁵ As a measure of gender equality we used the GII, and as a measure of economic development the HDI as provided by the Human Development Report (United Nations Development Programme, 2013). Due to the very high correlations between these four cultural dimensions (up to $r = -.85$ between HDI and GII) and the resulting problems of multicollinearity in the multivariate analyses, we used a composite index of emancipation as a mean value of all four cultural dimensions after standardizing all dimensions and recoding GII and PDI. The Scree-test of a factor analysis showed a clear one-factor solution. The factor explained 72.96% of the variance.

Page 9

Grau et al.

345

The emancipation index had a Cronbach's $\alpha = .88$. The CR index at the country level was calculated by averaging the values from all respondents in each country (Cronbach's $\alpha = .85$). Finally, aggregated measures of the response style facets and the response style index were computed for all countries (Cronbach's α for the response style index was .58).

Analytical Procedure

To assess the structure of CR, a principal component analysis of the five CR facets was con-

Analytical Procedure

To assess the structure of CR, a principal component analysis of the five CR facets was conducted at individual and country level. Then, individual level CR was regressed in a multilevel analysis on age, sex, education level, and the Big 5 personality traits (Nezlek, Schröder-Abé, & Schütz, 2006). In addition, the aggregate emancipation index based on the four cultural dimensions, HDI, GII, IDV, and PDI, was entered in the multilevel analysis.

All variables from Level 1 (e.g., age, sex, education level, Big 5 traits) were centered around their group means (e.g., mean of the country); the variable from Level 2 (e.g., emancipation) was centered around the grand mean. This analysis was conducted with HLM7 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). In the hierarchical linear modeling (HLM) analysis, random intercepts and fixed slopes were employed.

In the next step, the CR facets and the CR Index were treated as variables at the country level. Correlations with the cultural dimensions were computed. These correlations were also computed with response styles. Finally, we compared CR and response styles by computing correlations and a common factor analysis at the individual and at the country level.⁶

Results

Factor Structure of CR

A principal component analysis with Varimax rotation (oblimin rotation yielded virtually the same results) of the five CR indices at the individual level was conducted. The Scree-test suggested the extraction of one or two factors. Eigenvalues were 1.70, 1.06, 0.86, 0.76, and 0.63. In the one-factor solution, all factor loadings were above 0.50 except longstring (see Table 2). This result replicates the results of Meade and Craig (2012) and Huang et al. (2012). The authors found that longstring loaded on a different factor than the other facets of CR. In the two-factor solution, longstring loaded highly on the second factor. A problem for interpreting the two-factor solution was that psychometric antonyms and person-total correlation loaded on both factors. Therefore, it was decided to compute one composite CR index including all five facets after

Factor Structure of CR

A principal component analysis with Varimax rotation (oblimin rotation yielded virtually the same results) of the five CR indices at the individual level was conducted. The Scree-test suggested the extraction of one or two factors. Eigenvalues were 1.70, 1.06, 0.86, 0.76, and 0.63. In the one-factor solution, all factor loadings were above 0.50 except longstring (see Table 2). This result replicates the results of Meade and Craig (2012) and Huang et al. (2012). The authors found that longstring loaded on a different factor than the other facets of CR. In the two-factor solution, longstring loaded highly on the second factor. A problem for interpreting the two-factor solution was that psychometric antonyms and person-total correlation loaded on both factors. Therefore, it was decided to compute one composite CR index including all five facets after standardization and after recoding psychometric synonyms and person-total correlation (Cronbach's $\alpha = .50$).

Table 2 shows the factor loadings of the five CR indices at the country level. In this analysis, a clear one-factor solution with only one eigenvalue greater than 1 (3.19) was found, the factor explained 63.75% of the variance. The composite index at the country level reached a reliability of Cronbach's $\alpha = .85$. In a two-factor solution, longstring loaded on the second factor only, and psychometric antonyms and person-total correlation again showed double loadings. Thus, the factor structure at the country level provided a good replication of the analysis at the individual level. Prokrustes rotation of both two-factor solutions (R. Fischer, 2012) showed a good structural equivalence between the individual and the country level (Tucker's Phi = .98 for the first factor and .97 for the second factor, respectively).

The distribution of the composite CR index at the individual level is shown in Figure 1. Due to standardization of the five CR indices, the mean value of the composite index was zero ($SD = .58$). The lowest value was about -1. Values greater than 1 were found for 469 respondents (5.6%). For identifying persons with high levels of CR, this cut-off point would be a candidate

Correlates of CR at the Individual Level

To predict the CR index, we fitted a multilevel model with the individual variables entered at Level 1 (sex, age, education level, and Big 5 personality traits), and the emancipation index at Level 2. The model is presented in Table 3. Individual CR was not predicted by age or sex, but the CR-score was lower in participants with higher education levels. Except openness, all Big 5 personality traits predicted CR that was more pronounced in more neurotic participants, and less in individuals scoring high in extraversion, conscientiousness, and agreeableness. These results confirm Hypothesis 1 and replicate the findings of Bowling et al. (2016), although they used only respondents from the United States and a somewhat different CR composite index (without long-string). In addition, higher emancipation was a significant predictor of less CR too. Living in a more emancipated country was related to more carefully filling in a questionnaire.

Correlates of CR at the Country Level

Table 1 shows the mean values of CR for all participating countries. Participants from New Zealand showed the lowest level of CR (-0.37), while participants from Ecuador showed the highest mean CR-score (0.54). To explain these differences, we examined the association of CR facets and cultural variables. Table 4 shows the correlations between the five CR facets (long-string, Mahalanobis distance, psychometric synonyms, psychometric antonyms, person-total correlation, and the composite CR Index) with cultural dimensions. All facets of CR correlated negatively with the HDI and IDV, and positively with the GII and the PDI. Sixteen out of 20 of these correlations were significant. The highest correlations were found between the Mahalanobis distance and HDI ($r = -.81, p < .001$) and GII ($r = .85, p < .001$). The composite CR index and the composite emancipation index correlated ($r = -.76, p < .001$). The degree of emancipation of a country predicted the amount of CR in this country very well. The linear relationship between these two indices is presented in Figure 2. These results confirm the first part of Hypothesis 2.

The correlations between the response styles (except social desirability) and the cultural dimensions were also negative, but somewhat weaker than those with CR. The tendency to give

Correlates of CR at the Country Level

Table 1 shows the mean values of CR for all participating countries. Participants from New Zealand showed the lowest level of CR (-0.37), while participants from Ecuador showed the highest mean CR-score (0.54). To explain these differences, we examined the association of CR facets and cultural variables. Table 4 shows the correlations between the five CR facets (long-string, Mahalanobis distance, psychometric synonyms, psychometric antonyms, person-total correlation, and the composite CR Index) with cultural dimensions. All facets of CR correlated negatively with the HDI and IDV, and positively with the GII and the PDI. Sixteen out of 20 of these correlations were significant. The highest correlations were found between the Mahalanobis distance and HDI ($r = -.81, p < .001$) and GII ($r = .85, p < .001$). The composite CR index and the composite emancipation index correlated ($r = -.76, p < .001$). The degree of emancipation of a country predicted the amount of CR in this country very well. The linear relationship between these two indices is presented in Figure 2. These results confirm the first part of Hypothesis 2.

The correlations between the response styles (except social desirability) and the cultural dimensions were also negative, but somewhat weaker than those with CR. The tendency to give extreme answers correlated with all cultural dimensions, while avoiding midpoint answers correlated only with HDI and GII (see Table 4). The correlations between acquiescence and cultural dimensions were not significant. All correlations between social desirability and cultural

Comparison Between CR and Response Styles

The correlations between the CR facets and the response styles are shown in Table 5. It is noticeable that the response styles at the individual level showed higher intercorrelations between each other than the CR facets which leads to a higher reliability of the response style index compared with the CR index. The highest correlation was found between extreme and midpoint answers

Page 12

348

Journal of Cross-Cultural Psychology 50(3)

Table 4. Correlations of Country-Level Variables.

	HDI	GII	IDV	PDI	Emancipation index
Longstring	-.31	.34	-.58***	.44**	-.48**
Mahalanobis d	-.81***	.85***	-.56**	.45**	-.78***
Synonyms	-.53**	.51**	-.61***	.45**	-.61***
Antonyms	-.41*	.40*	-.69***	.48**	-.58***
Person-total correlation	-.29	.29	-.59***	.45**	-.47**
CR index	-.62***	.63***	-.76***	.57***	-.76***
Extreme	-.68***	.75***	-.40*	.35*	-.64***

Discussion

In this cross-cultural study on CR, it was shown that CR at the individual level correlated with education level and personality traits. This part of the present study replicated the findings of Bowling et al. (2016) with a nonstudent sample from 34 countries. In the present study, CR was examined both at the individual and country levels. The results showed that CR at the country level was strongly correlated with the emancipation level of countries. CR showed some overlap with response styles but loaded on its own factor and is more connected with cultural dimensions than response styles.

CR at the Individual Level

At the individual level, CR seems to be a construct that is relatively stable for more than a year (Bowling et al., 2016). In our total sample across 34 countries, the internal consistency was $\alpha = .50$ at the individual level, which is considerably lower than the .79 found by Bowling et al. (2016, Study 4). However, Bowling's data were based on a much more homogeneous U.S. student sample.

We found that all five CR facets were loading on one factor. In a two-factor solution, the facet longstring loaded alone on the second factor, replicating the finding of Meade and Craig (2012). These authors found three factors: one consisted mainly of facets used in our study, the second was longstring, and the third was self-reported effort in the study, a variable we did not include. It is obvious that longstring reflects a different answer behavior than psychometric synonyms, antonyms, Mahalanobis distance, and person-total correlation. Longstring answer patterns refer to a total lack of differentiation between the items. The other CR facets indicate that persons fail to respond correspondingly to similar content and to discriminate between different content, and that they overlook reverse-coded items. This behavior can result in a large Mahalanobis distance and a low person-total correlation. For statistical reasons (double loadings of two variables), we decided to compute only one composite index of CR. However, inclusion of more CR facets (e.g., self-reported effort) may lead to describing CR as a multidimensional construct including

CR at the Individual Level

At the individual level, CR seems to be a construct that is relatively stable for more than a year (Bowling et al., 2016). In our total sample across 34 countries, the internal consistency was $\alpha = .50$ at the individual level, which is considerably lower than the .79 found by Bowling et al. (2016, Study 4). However, Bowling's data were based on a much more homogeneous U.S. student sample.

We found that all five CR facets were loading on one factor. In a two-factor solution, the facet longstring loaded alone on the second factor, replicating the finding of Meade and Craig (2012). These authors found three factors: one consisted mainly of facets used in our study, the second was longstring, and the third was self-reported effort in the study, a variable we did not include. It is obvious that longstring reflects a different answer behavior than psychometric synonyms, antonyms, Mahalanobis distance, and person-total correlation. Longstring answer patterns refer to a total lack of differentiation between the items. The other CR facets indicate that persons fail to respond correspondingly to similar content and to discriminate between different content, and that they overlook reverse-coded items. This behavior can result in a large Mahalanobis distance and a low person-total correlation. For statistical reasons (double loadings of two variables), we decided to compute only one composite index of CR. However, inclusion of more CR facets (e.g., self-reported effort) may lead to describing CR as a multidimensional construct including different factors that may have different correlates and consequences. This should be investigated in future research.

Although using slightly different CR facets, we fully replicated the correlations between CR and the Big 5 personality traits found by Bowling et al. (2016). In the current study, the Big 5 were self-reported, while in the study by Bowling et al., they were reported by acquaintances. In both studies, CR correlated negatively with conscientiousness, agreeableness, extraversion, and emotional stability. In addition, the emancipation index correlated negatively with CR, showing that the emancipation index of countries can explain individual differences in CR.

participate in the study than those in the United States, Moldova, or New Zealand. Rather, we assume that cultures systematically differ in familiarity with questionnaires or in appreciation of psychological research. One could assume that the educational level differs between countries. Indeed, there was no correlation between country-level CR and country-level formal education ($r = -.03$, $p = .876$), but it is possible that, for example, a university degree means different things in the 34 participating countries. Another possible explanation for cultural differences in CR might be found in the different participation rates. It was not assessed how many people had to be addressed to distribute the questionnaires and how many declined to take part in the survey (which would be something different than return rates). It is conceivable that individuals in more emancipated, individualistic countries, with lower power distance who had difficulties handling questionnaires or who were not motivated to do so, felt free to refuse to participate, whereas

questionnaires or who were not motivated to do so, felt free to refuse to participate, whereas participants in more collectivistic countries with higher power distance tended not to do that.

Correspondingly, Smith (2004) found a higher level of an acquiescent response style in collectivistic cultures and in countries with high levels of power distance. As response styles reflect not only a behavior in completing questionnaires but also a broader communication style, it is possible that persons in these countries also tended to answer "yes" more often if asked to participate in a survey although they were not very motivated to do so.

It is a surprising result that CR at the country level correlates extremely high with indicators of cultural emancipation. Country-level CR is more reliable than individual level CR (which is not the case for response styles). Studying CR at the country level could be more interesting than CR on the individual level. It can be considered as a powerful predictor of intercultural differences itself, and as a mediator variable which explains why participants in countries with higher or lower levels of emancipation, human development, or individualism answer questions differently. We suggest to measure the effects of country level CR in cross-cultural studies. Only a short questionnaire with a few constructs and some reverse-coded items is needed for assessing CR. In contrast to He and van de Vijver (2015), we used the same items to compute all CR and response style facets, and in contrast to Meade and Craig (2012), we used only a few item pairs to compute synonyms and antonyms. Nevertheless, the meaningful factor structure and the high reliability of the country-level CR show that even a small number of items can be used to assess CR. As a very easy measure of CR at the country level, the reliability score of a scale or the item-total correlation of a reverse-coded item could also be used as a proxy. We do not suggest partialling out effects of CR from interesting correlations with cultural dimensions such as HDI, IDV, GII, and PDI due to the high overlap between CR and cultural dimensions. He, Bartram, et al. (2014) studied response styles at the country level and argued that "response styles are better not corrected for in the measurement" (p. 1042). "If we were to correct for such response styles, we would eliminate valid personality variance" (He, Bartram, et al., 2014, p. 1041; cf. R. Fischer, 2004). In this respect, CR is structurally similar to response styles. Instead of controlling for CR, this variable can be treated as an additional cultural dimension or as a mediator between cultural dimensions and individual responses, and the causal relationships between CR and other cultural dimensions should be investigated in further research.

normally seen as an effect of motivation to present oneself in a special way, whereas CR is seen as a result of a low motivation to answer questions. In contrast, social desirability is the only one of all response styles that requires careful reading of all items. Giving midpoint or extreme answers or saying “yes” independently from the content of the items can be done without reading any items. Therefore, we expected positive correlations between CR and all response styles except social desirability and found this pattern fully confirmed at the individual level and partly confirmed at the country level.

Two correlations between CR and response styles were especially high. The Melanesian

Two correlations between CR and response styles were especially high: The Mahalanobis distance correlated with extreme answers. If one tends to prefer answering with "1" or "5" on a 5-point Likert-type scale, one also tends to be an outlier. If one answers in a socially desirable manner, one also tends to show an answer profile that is equal to the profile of all other participants. These overlaps between CR and response styles are comprehensible, but nevertheless CR and response styles are distinct constructs which load on different factors.

How to Deal With CR at the Country Level?

Systematically different amounts of CR between countries can imply misleading conclusions about cultural differences. The exclusion of participants showing high levels of CR does not solve that problem. If we exclude all persons with more than two standard deviations above the CR-mean (4%), the correlations between CR and personality traits, education level, and emancipation at the individual level remain constant (Table 3). Also, the correlation between CR and the emancipation index at the country level remains ($r = -.76$; Table 4). There are several suggestions to deal with cultural differences in response styles so far. For example, it is well documented that in PISA-studies (Programme for International Student Assessment), assessment scores and self-reported ability or motivational scores were often negatively correlated at the country level (Buckley, 2009; Vonkova, Papajoanu, & Stipek, 2018). This can be interpreted as a result of different scale use in the participating countries. One solution of this problem is the implementation of the overclaiming technique (Vonkova et al., 2018). In this approach, not only the familiarity of existing concepts is asked for but also the familiarity with nonexisting concepts. If participants in some countries report to know nonexisting facts, these reports can be used to adjust the self-reported knowledge. A second technique is the implementation of anchoring vignettes in self-reports (He, Buchholz, & Klieme, 2017; Kyllonen & Bertling, 2013). This means that concrete attributes of some vignettes are described. The self-ratings can be rescaled using the vignette ratings as anchors. However, these techniques assume that mean values are systematically distorted between countries. This is not the case with CR. The tendency to say yes or no can be more easily corrected than the tendency to overlook reverse-coded items because this error has more impact on correlations, factor structures, and reliability than on mean values.

How to Deal With CR at the Country Level?

Systematically different amounts of CR between countries can imply misleading conclusions about cultural differences. The exclusion of participants showing high levels of CR does not solve that problem. If we exclude all persons with more than two standard deviations above the CR-mean (4%), the correlations between CR and personality traits, education level, and emancipation at the individual level remain constant (Table 3). Also, the correlation between CR and the emancipation index at the country level remains ($r = -.76$; Table 4). There are several suggestions to deal with cultural differences in response styles so far. For example, it is well documented that in PISA-studies (Programme for International Student Assessment), assessment scores and self-reported ability or motivational scores were often negatively correlated at the country level (Buckley, 2009; Vonkova, Papajoanu, & Stipek, 2018). This can be interpreted as a result of different scale use in the participating countries. One solution of this problem is the implementation of the overclaiming technique (Vonkova et al., 2018). In this approach, not only the familiarity of existing concepts is asked for but also the familiarity with nonexisting concepts. If participants in some countries report to know nonexisting facts, these reports can be used to adjust the self-reported knowledge. A second technique is the implementation of anchoring vignettes in self-reports (He, Buchholz, & Klieme, 2017; Kyllonen & Bertling, 2013). This means that concrete attributes of some vignettes are described. The self-ratings can be rescaled using the vignette ratings as anchors. However, these techniques assume that mean values are systematically distorted between countries. This is not the case with CR. The tendency to say yes or no can be more easily corrected than the tendency to overlook reverse-coded items because this error has more impact on correlations, factor structures, and reliability than on mean values. Buckley's (2009) results lead to the conclusion that the tendency to say yes or no can be corrected for in cross-cultural studies but that CR only adds noise to the measurement. Nevertheless, it should be explored in future studies, whether CR can be diminished by asking more concrete questions or using forced choice answering formats (Kyllonen & Bertling, 2013).

Limitations

Limitations

The current study included a big sample from more than 30 countries. Unlike many other studies, not only students (of psychology) were recruited but a more diverse sample with a wide range of age and educational levels took part in the survey. Nevertheless, some limitations of the current study have to be noted. First, the Mahalanobis distance and the person-total correlation were calculated over the whole sample and not in each country separately to ensure that cultural differences are not eliminated. Being an outlier compared with more than 8,000 persons from 34

Page 18

countries is somewhat different from being an outlier in a data set from one's own country. It could be argued that persons in some countries are systematic outliers because the translation of the questionnaire was not perfect and the items in one country measured something different than in other countries. We cannot entirely exclude this alternative explanation, although the translation and back-translation process was conducted very carefully. However, there are strong arguments to challenge this alternative explanation. First, the amount of CR was very different even in countries with the same language (e.g., Ecuador and Spain). Second, we precisely replicated the correlation patterns between CR and the Big 5 from Bowling et al. (2016) who used question-

in countries with the same language (e.g., Ecuador and Spain). Second, we precisely replicated the correlation patterns between CR and the Big 5 from Bowling et al. (2016) who used questionnaires in only one language in one country. Third and most importantly, CR at the country level showed high correlations with other cultural dimensions. If the differences between countries were only measurement errors, they could not be explained by other variables.

A second point that has to be noted is that we found the longstring response style to load on a second factor but included it in the CR composite index. Answering repeatedly with the same response is psychologically probably somewhat different than overlooking reverse-coded items. In the study by Meade and Craig (2012), longstring loaded on a different factor, but the authors used more than one longstring indicator, which makes it easier to find an extra factor. Future research should explore whether there are additional aspects of CR, and whether CR should be conceived of as a mono- or multidimensional construct with different external correlations of the subdimensions. It is also noteworthy that not all indicators of CR were included in the current study; for example, response times or Guttman errors. Finally, we did not include bogus items and instructed items in the study. In the study of Meade and Craig (2012), bogus items loaded on the same factor as other facets that were included in the current study. Unfortunately, little is known about the influence of instructed items yet, items like "please leave this item blank" or "please answer here with I totally agree." In a pilot study with 121 psychology students, we used three of these items and found that some participants intentionally failed to answer these questions correctly, for example, by purposefully checking an incorrect answer and commenting with a smiley or something like "haha." More importantly, the percentage of correctly answered instructed items did not correlate with any other CR facet or with any construct of interest. These first results suggest that at least for persons who are very familiar with questionnaires (e.g., psychology students), instructed items have to be used with caution. Due to their high transparency, instructed items can lead to reactance, that is, answering wrongly on purpose. For a similar reason, Huang et al. (2012) also excluded these items from data analysis.

Conclusion

Taken together, the results of the present study indicate that CR at the individual level is a



Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Grant BA 1603/4-1 and 4-2 of the German Research Foundation (DFG) to R. Banse.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In Latvia and Moldova, the Russian version of the questionnaire was used, which is why only Russian-speaking participants could be recruited.
2. The wording of the Big 5 items is shown in Supplemental Table 1.
3. Means and standard deviations of the Big 5 personality traits are presented in the Supplemental Table 2.
4. Means, standard deviations, and ranges of careless responding indices and response styles can be found in the Supplemental Table 3.
5. Unfortunately, for Moldova and Georgia no data regarding the Hofstede dimensions are available. Due to a large cultural overlap based on the common history until the dissolution of the Soviet Union, we used the Russian data instead.
6. Questionnaires, data, and SPSS-syntax can be obtained from the first author.

References

- Ahammer, I. M. (1971). Desirability judgments as a function of item content, instruction set, and sex: A life-span developmental study. *Human Development*, 14, 195-207. doi:10.1159/000271213
- Archer, J. (2006). Cross-cultural differences in physical aggression between partners: A social-role analysis.

References

- Ahammer, I. M. (1971). Desirability judgments as a function of item content, instruction set, and sex: A life-span developmental study. *Human Development*, 14, 195-207. doi:10.1159/000271213
- Archer, J. (2006). Cross-cultural differences in physical aggression between partners: A social-role analysis. *Personality and Social Psychology Review*, 10, 133-153. doi:10.1207/s15327957pspr1002_3
- Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the big five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729-750. doi:10.1037//0022-3514.75.3.729
- Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, 111, 218-229. doi:10.1037/pspp0000085
- Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Methodology: Vol. 2. Handbook of cross-cultural psychology* (pp. 389-444). Boston, MA: Allyn & Bacon.
- Buckley, J. (2009). *Cross-national response styles in international educational assessments: Evidence from PISA 2006*. Department of Humanities and Social Sciences in the Professions, New York University. Retrieved from https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 60, 151-174. doi:10.1037/h0040372
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475-494. doi:10.1177/001316444600600405
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4-19. doi:10.1016/j.jesp.2015.07.006
- Desimone, J. A., Harms, P. D., & DeSimone, A. J. (2014). Best practice recommendations for data screening. *Journal of Organizational Behavior*, 36, 171-181. doi:10.1002/job.1962

- Dillon, M. (2014). *Introduction to sociological theory: Theorists, concepts, and their applicability to the twenty-first century* (2nd ed.). Chichester, UK: Wiley-Blackwell.
- Ebbeler, C., Grau, I., & Banse, R. (2017). Cultural and individual factors determine physical aggression between married partners: Evidence from 34 countries. *Journal of Cross-Cultural Psychology*, 48, 1098-1118. doi:10.1177/0022022117719497
- Eysenck, S. B. G., & Eysenck, H. J. (1963). An experimental investigation of "desirability" response set in a personality questionnaire. *Life Science*, 2, 343-355. doi:10.1016/0024-3205(63)90168-1
- Fischer, A. H., & Manstead, A. S. R. (2000). The relation between gender and emotions in different cultures. In A. Fischer (Ed.), *Studies in Emotion and Social Interaction. Gender and emotion: Social psychological perspectives* (pp. 71-94). Cambridge, UK: Cambridge University Press.
- Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, 35, 263-282. doi:10.1177/0022022104264122
- Fischer, R. (2012). Value isomorphism in the European Social Survey: Exploration of meaning shifts in values across levels. *Journal of Cross-Cultural Psychology*, 43, 883-898. doi:10.1177/0022022111413276
- Hampson, S. E. (1998). When is an inconsistency not an inconsistency? Trait reconciliation in personality description and impression formation. *Journal of Personality and Social Psychology*, 74, 102-117. doi:10.1037/0022-3514.74.1.102
- He, J., Bartram, D., Inceoglu, I., & Van de Vijver, F. J. R. (2014). Response styles and personality traits: A multi-level analysis. *Journal of Cross-Cultural Psychology*, 45, 1028-1045. doi:10.1177/0022022114534773
- He, J., Buchholz, J., & Klieme, E. (2017). Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures. *Journal of Cross-Cultural Psychology*, 48, 319-334.
- He, J., & Van de Vijver, F. J. R. (2015). Effects of a general response style on cross-cultural comparisons: Evidence from the Teaching and Learning International Survey. *Public Opinion Quarterly*, 79, 267-290. doi:10.1093/poq/nfv006
- He, J., Van de Vijver, F. J. R., Dominguez Espinosa, A., & Mui, P. H. C. (2014). Toward a unification of acquiescent, extreme, and midpoint response styles: A multilevel study. *International Journal of Cross Cultural Management*, 14, 306-322. doi:10.1177/1470595814541424
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.