Original Article

Efficient Measurement of Global Human Identification and Citizenship – Is One Item Sufficient?

Philipp Geisen, Friedemann Trutzenberg, and Michael Eid

Department of Education and Psychology, Freie Universität Berlin, Germany

Abstract: To address the complexity of global challenges (e.g., migration, climate change), research must examine many constructs and their interrelationships. Therefore, efficient assessment methods (e.g., short scales) are required, especially when constructing large-scale social surveys. A widely used instrument for assessing global identification as an important predictor of behavior related to coping with global challenges is the Identification With All Humanity (IWAH) scale (McFarland et al., 2012). With the present study, we aimed to investigate whether this scale can be sufficiently represented by a single item. Using bifactor-(S*I − 1) models, we analyzed whether one item from the scale itself (Item 5) and a related single-item measure from the World Values Survey (WVS; Inglehart et al., 2018) could adequately represent a German version of the IWAH scale. The sample (*n* = 1,369) was representative of the German population regarding age, gender, income, education, and region of residence. Item 5 showed an appropriate single-item reliability (.82) and represented 64–93% of the true variance in the individual IWAH items, whereas the WVS item (Rel(Y_{WVS}) = .61) represented 52–81%. To reflect the multidimensionality of the IWAH, a 4-item short scale is recommended.

Keywords: bifactor-(S*I − 1) model, global human identification and citizenship, identification with all humanity, IWAH scale, short scale

Over the last decade, complex interdependencies of people around the world have become clear from events such as the rapid spread of COVID-19 and climate change. These challenges can be met only through global cooperation. As identification with a group is important for acting in favor of its welfare (Reese et al., 2015), research on *global human identification and citizenship* (GHIC) has increased in recent years (McFarland et al., 2019). Whereas global human identification means that an individual identifies with all human beings, global citizenship refers to considering oneself as "belonging to the global collection of human beings" (McFarland et al., 2019, p. 142). As the two facets are closely related, GHIC is considered a singular construct (McFarland et al., 2019). According to a recent review of empirical psychological studies by McFarland et al. (2019), GHIC is correlated with lower prejudice and a greater willingness to accept members of outgroups as fellow citizens. Additionally, it predicts greater concern for human rights or the global environment, a greater (desire for) global knowledge, and higher donations to international charities. These results demonstrate the importance of GHIC for research on current global challenges.

A widely used scale for assessing global human identification is the *Identification With All Humanity* (IWAH) scale (McFarland et al., 2012), which consists of nine items. Respondents are asked to rate their agreement with nine statements on a 5-point unipolar rating scale with respect to people (a) in their own community, (b) of their nation, or (c) all over the world (i.e., humanity). For further analyses, the sum value of each group is considered a measure of the respective kind of identification. The internal consistencies of the three subscales are satisfactory (α > .8; McFarland et al., 2012). Regarding construct validity, the factor structure of the IWAH scale is controversial. Whereas McFarland et al.'s (2012) exploratory factor analyses suggested that all nine items loaded on one general factor, other studies supported a two-factor solution (e.g., McFarland & Hornsby, 2015). According to this approach, Items 1–4 load on one factor and Items 6–9 on a second one; Item 5 should be excluded because of its high loadings on both factors. As the factors are strongly correlated, a hierarchical model with one second-order factor also fits well (Hamer et al., 2021).

Among the supporters of the two-factor solution, there is a discussion about its interpretation: Referring to a

https://doi.org/10.1027/1015-5759/a000801

multi-component model of ingroup identification (Leach et al., 2008), Reese et al. (2015) postulated that the first factor (*global self-definition*) concerns the individual self-categorization as a group member, whereas the second (*global self-investment*) reflects the behavioral and emotional involvement induced by group membership. According to Reysen and Hackett (2016), the first factor reflects the social identity approach and the second factor the perspectives of theories by differential psychologists, such as Adler and Maslow.

Hamer et al. (2021) criticized both proposals by claiming that the items from the IWAH scale do not match the components of ingroup identification specified by Leach et al. (2008) and Reysen and Hackett's (2016) approach is too vague. Hamer et al. interpreted the first factor as a bond with all humanity through self-categorization and affinity and the second factor as proactive caring for other group members, including feelings of loyalty and responsibility.

The IWAH scale measures the construct in detail but might be too long for studies of complex (psychological or social) phenomena in which relationships between many different constructs are examined. In particular, large studies often require short scales to avoid boredom, fatigue, and motivation impairments that might result in lower data quality and a reduction in the response rate (Rammstedt & Beierlein, 2014). For such large studies, single-item measures as a kind of short scale have crucial advantages regarding efficiency (e.g., reduction in data processing costs) but also validity: Despite lower content validity for the measurement of complex constructs, single-item measures are especially useful for constructs that are clearly defined and narrow in scope (e.g., GHIC), as they can have higher face validity than multi-item measures that include unclear items (Allen et al., 2022; Rammstedt & Beierlein, 2014). The present study aimed to analyze whether a single item could provide a sufficient assessment of GHIC. First, we examined whether the IWAH scale could be sufficiently represented by a single item from the scale itself. Because it has high loadings on both factors and the largest item-total correlation (McFarland et al., 2019), Item 5 ("How much do you identify with [that is, feel a part of, feel love toward, have concern for] each of the following [groups]?"; McFarland et al., 2012, p. 852) was chosen as the most appropriate candidate. Second, single items are already used in large-scale social surveys such as the World Values Survey (WVS; Inglehart et al., 2018) to measure GHIC. The question of the degree to which the WVS item ("I see myself as a world citizen"; Inglehart et al., 2018, WV6 Questionnaire, V212) represents the more complex IWAH scale is also important for relating the research done with the WVS to research on GHIC. Finally, it is important to analyze whether short scales or single items show test-criterion correlations that are similar to those found for long forms (Ziegler et al., 2014).

Method

We report how we determined our sample size, all data exclusions, all data exclusion criteria, whether exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact *p*-values, effect sizes, and 95% confidence or credible intervals.

Participants and Procedures

The data used in this article stemmed from a larger study that focused primarily on analyzing perceived parenting and IWAH with structural equation modeling (Hagel et al., 2022). The sample size was determined a priori with the Shiny App *pwrSEM* (Wang & Rhemtulla, 2021). This sample size is also appropriate for our analyses for two reasons. First, the models presented in this paper were smaller (with respect to the number of observed and latent variables) than the model from the a priori power analysis. Second, simulation studies for analyzing categorical data with structural equation modeling (by using the WLSMV estimator) indicated that the sample size was sufficient (Nussbeck et al., 2006).

Data were collected via an online survey in June and July 2021. Participants were recruited from English and German panels provided by the company *respondi*. To guarantee that each subsample was representative of the corresponding population with respect to five variables (gender, age, region of residence, education, and income), the company invited 2,998 members of the German and 3,593 of the English panel to fill out the main questionnaire. They were chosen to meet population quotas on the basis of their answers on an initial screening questionnaire. The main questionnaire contained three quality check items to detect whether participants had worked thoroughly (e.g., "Please select option 1 so that we can conduct data quality checks"). Participants were excluded from our analyses when they had (a) not agreed to the terms of participation, (b) not completed the questionnaire through the last page, or (c) not answered any item that was relevant to our analyses. Besides, participants were excluded (d) if they had not answered the first and/or third quality check correctly, and (e) if their data contained (more than) twice as many other indications of poor data quality than the mean number in the sample (per completed survey page). These indications were implausible or unrelated answers on certain questions, inconsistencies between the screening and the main questionnaire or between answers on the main questionnaire, speeding, or straightlining (see Bowling, 2005; Zhang & Conrad, 2014). As the second quality check item had been positioned on a survey page that was relevant only for a subsample, it did not qualify for quality checks for the complete sample. Applying these five criteria led to the exclusion of 3,888 participants and a final sample size of N = 2,703 (English: n = 1,334; German: n = 1,369).

Depending on the country of recruitment, each participant received an English or German version of the survey. Participants answered a large set of questionnaires that measured their global identity, personality, and well-being (among other constructs). A correlational, cross-sectional research design was employed.

After the data were collected, we realized that, for a subgroup of individuals, the wording of the English versions of two measures that were relevant for the present study (IWAH scale, WVS item) differed slightly from the original wording by accident. As it was unclear how these differences might influence the items' psychometric properties, the analyses were conducted on only the German data.

The German subsample (M_{age} = 46.68 years) consisted of 49% females, 49% males, and 0.22% nonbinary participants. The median monthly income was 3,100 €. A total of 95% of the participants had only German nationality, but about 46% spoke at least one second language besides German fluently (mainly English). Additionally, 93% had already been abroad.

Measures

Identification With All Humanity Scale

The German translation of the IWAH scale (Reese et al., 2015) was adapted: Whereas respondents who categorized themselves as members of a religious community received a five-part version (assessing the identification with people in their own community, German people, Europeans, adherents of their own religion, people all over the world), nonmembers received a four-part one (missing the reference to the religious group). Additionally, the verbal anchors of the 5-point unipolar rating scale (for most items: *not at all*, *just a little*, *somewhat*, *quite a bit*, *very much*) were presented with letters ranging from "A" to "E" instead of numerical anchors. We focus only on the answers to the part of the scale that refers to all of humanity. As the distinction between members and nonmembers of religious communities was not relevant to this study, these groups were analyzed together. The items of the IWAH scale are shown in the Appendix.

World Values Survey Item

Various values, beliefs, and attitudes of people from 120 countries and their development have been investigated by the WVS since 1981 every 5 years. For example, in the German version of Wave 6 (2010–2014; Inglehart et al., 2018), participants were asked to rate the single item regarding global citizenship cited above on a 4-point unipolar rating scale with verbal (*strongly disagree, somewhat disagree, somewhat agree, strongly agree*) and numerical anchors. The online survey was adapted slightly to present the term "world citizen" in a way that was not gender-biased in German ("Weltbürger*in").

Further Measures

To examine the test-criterion correlations for the single items as well as for the total score of the long form, the following constructs from the main questionnaire were considered: Well-being was measured with the Comprehensive Inventory of Thriving (Su et al., 2014); personality with the 10-item version of the Big Five Inventory (Rammstedt et al., 2014); mothers' and fathers' parenting styles with a combination of items from the Measure of Parental Style (Rumpold et al., 2002), the Zurich Brief Questionnaire for the Assessment of Parental Behaviors (Reitzle et al., 2001), a translated version of the Evaluation of Parental Educational Practices (Meunier & Roskam, 2007), and four newly developed items.

Outline of the Main Analyses

To address the main questions, *bifactor-($S^*I - 1$) models* (Eid et al., 2017, 2018), structural equation models for analyzing *G*-factor structures, were applied. In this kind of bifactor model, one item loads only on the general factor (G) but not on any specific factor (S_k; see Figures 1 and 2). This item, which is selected on the basis of theoretical considerations, defines the *G*-factor as its reference indicator (RI). All other items load on the *G*-factor and (at least) one specific factor. The observed variance for the nonreference indicators can be decomposed into three parts (Eid et al., 2017): (a) *consistency* as the percentage explained by the *G*-factor/shared with the RI; (b) *specificity* as the percentage explained by the specific factor(s); and (c) measurement error. Consistency and specificity add up to the reliability as the explained part of the variance. Additionally, consistency and specificity coefficients can be defined with respect to the true score variance of an observed variable, so that they do not depend on the reliabilities and can be better compared across the observed variables. In this case, they add up to 1. The variance of an RI can be decomposed into only the part explained by the *G*-factor and measurement error. Therefore, only the reliability coefficient (but no consistency or specificity) can be estimated. As the reliability coefficient does not depend on the number of items, this study can handle the important argument against short scales that their reliability is lower or unknown (Allen et al., 2022; Ziegler et al., 2014).

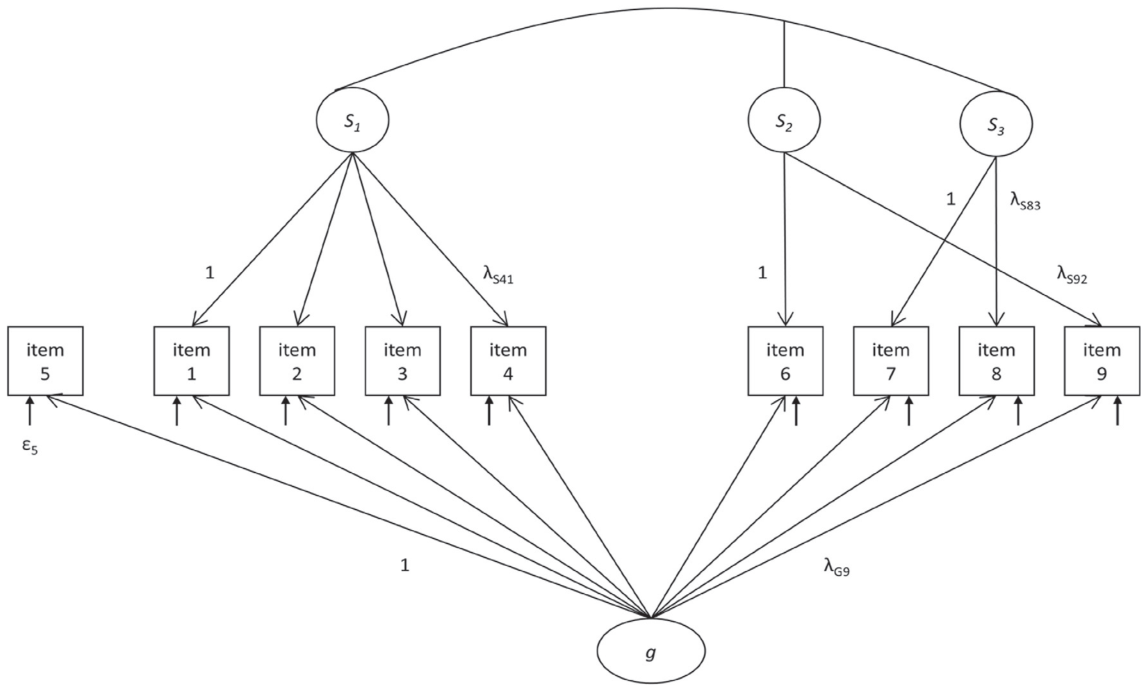

Figure 1. Modified Model A.

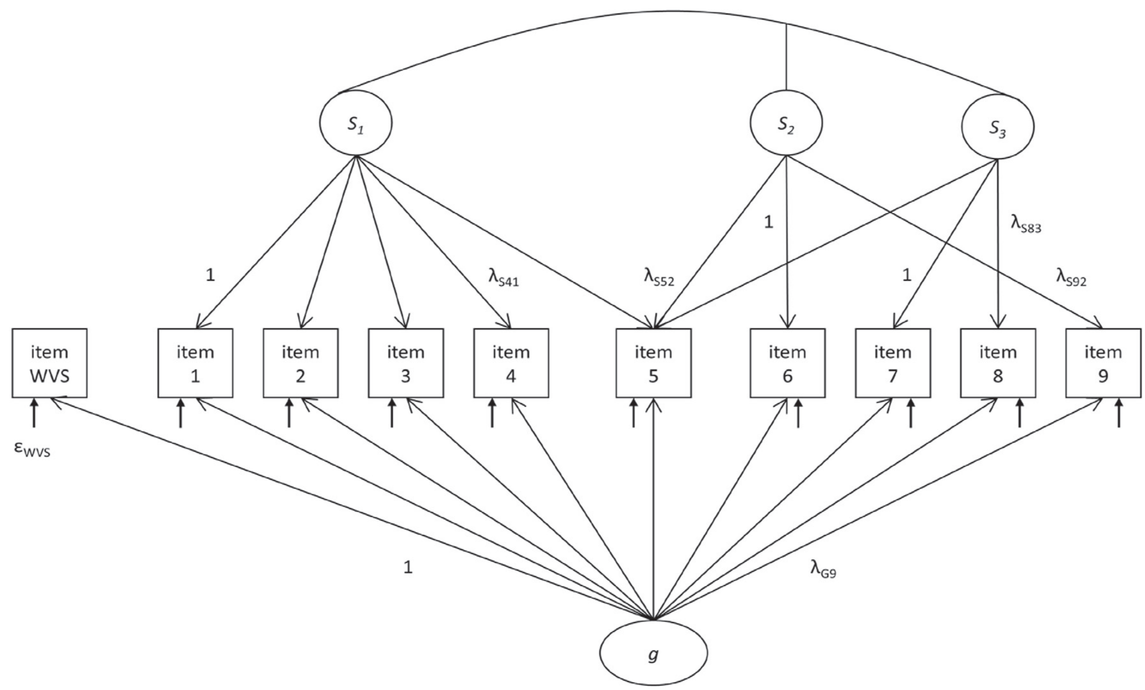

Figure 2. Modified Model B.

Two bifactor-(S*I − 1) models were investigated: In Model A, Item 5 from the IWAH scale served as the RI; in Model B, the WVS item did. According to previous research on the factor structure of the IWAH scale, the baseline versions of these models include two specific factors (S_1: Items 1–4; S_2: Items 6–9). Because of its high loadings on both factors found in previous studies, Item 5 was included as an indicator that loaded on each specific factor in Model B.

The bifactor-(S*I − 1) models were analyzed with the lavaan program (Rosseel, 2012), Version 0.6-11. Because the observed variables were single items with ordered

Table 1. Polychoric correlations between the individual items

Item	1	2	3	4	5	6	7	8	9	WVS
1	–	.70	.72	.60	.72	.60	.57	.54	.60	.57
2		–	.70	.63	.67	.53	.48	.46	.56	.51
3			–	.61	.72	.57	.60	.54	.59	.57
4				–	.65	.50	.39	.43	.52	.44
5					–	.69	.67	.61	.68	.58
6						–	.61	.61	.72	.50
7							–	.71	.70	.61
8								–	.70	.51
9									–	.55

Note. WVS = World Values Survey

response categories, we used a weighted least squares mean and variance (WLSMV) adjusted estimator, which is based on the polychoric correlation, to estimate the model parameters and fit. As the χ^2 test statistic depends on the sample size, which was quite large in the present study, model fit was evaluated by the following descriptive goodness-of-fit criteria according to Schermelleh-Engel et al. (2003): (a) root mean square error of approximation (RMSEA) $\leq$ .05; (b) comparative fit index (CFI) $\geq$ .97; and (c) standardized root mean square residual (SRMR) $\leq$ .05.

To address the research questions, the reliabilities and consistency coefficients of the indicators were of major interest. The higher the non-reference indicators' consistency coefficients, the more variance the RI explains in them. There is no general rule on how large consistency coefficients should be. Rather, the guidelines depend on the area of application and the available resources. Thus, we did not define a lower limit for the consistency coefficients or reliability a priori.

Results

The strength of the correlations was classified on the basis of Funder and Ozer (2019; very small = 0.05, small = .10, medium = .20, large = .30, very large = .40 or greater).

Descriptive Statistical Analyses

There were no missing values on the relevant variables. Table 1 presents the polychoric correlations between the individual items. In general, all correlations were (very) large (.39 or greater) and rather similar. However, Item 5 was the only item for which all correlations with the other IWAH items were larger than .60. Moreover, Table 2 presents the relative frequency distributions for the individual items. They showed that it was hardest to agree with Items 2 and 4 and easiest with Items 7 and 8. For all other items, the middle category 3 was the median and the mode.

Regarding the test-criterion correlations, Table 3 demonstrates that the correlations between several external constructs and the chosen reference indicators/their multi-item counterparts were quite similar.

Main Analyses

Regarding Model A, the RMSEA (RMSEA = .070, p < .01) indicated an insufficient fit of the baseline version, but the CFI (CFI = .994) and SRMR (SRMR = .019) indicated a very good fit (see Table 4). The modification indices indicated that a theoretically meaningful modification would be to split the second specific factor into two factors (S_2: Items 6 and 9; S_3: Items 7 and 8). This action was also theoretically justifiable, as Items 6 and 9 reflected a dimension of caring, whereas Items 7 and 8 measured the subjective importance of representing moral values. This modification (see Figure 1) fit the data well (RMSEA = .051, p_{RMSEA} = .399; CFI = .997, SRMR = .016). Because the same IWAH items were analyzed in Model B, we specified Model B according to this modification (see Figure 2) and did not evaluate the baseline model. The modified Model B fit the data very well (RMSEA = .055, p_{RMSEA} = .188; CFI = .996, SRMR = .016).

Tables 5 and 6 report the (un)standardized loading parameters as well as the reliability, consistency, and specificity coefficients for the modified Models A and B. In Model A, the standardized factor loadings, which can be interpreted as correlations between the factors and the observed variables, were very high for the *G*-factor (.677 to .905) and comparably low for the specific factors (.223 to .563). Consequently, the consistency coefficients of the non-reference indicators with respect to the true variance were very high (.638 to .929). Moreover, the reliabilities were rather high for the single-item measures (.544 to .875); the reliability of the RI, Item 5, was .819. The second and third specific factors were correlated very largely (.640); their correlations

Table 2. Relative frequency distributions

Response option	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item WVS
1	.10	.29	.13	*.49*	.19	.10	.11	.09	.10	.14
2	.24	.26	.25	.29	.27	.16	.13	.13	.14	.27
3	*.42*	*.29*	*.47*	.17	*.37*	*.36*	.28	.23	*.35*	*.45*
4	.20	.12	.13	.04	.13	.27	*.31*	*.31*	.26	.14
5	.04	.04	.02	.02	.03	.11	.17	.25	.15	–

Note. If a value is italics, the associated response option is the mode. Inconsistencies are due to rounding errors. WVS = World Values Survey.

Table 3. Criterion correlations of the single items and their multi-item counterpart

Construct	IWAH scale[a]	Item 5[b]	WVS Item[b]
Well-being	.34	.28	.33
Personality trait: openness	.27	.25	.28
Personality trait: conscientiousness	.14	.09	.09
Personality trait: extraversion	.15	.12	.17
Personality trait: agreeableness	.27	.24	.24
Personality trait: neuroticism	−.04	−.02	−.08
Parenting style of mother	.09	.09	.06
Parenting style of father	.07	.06	.04

Note. [a]Pearson correlation. [b]Polyserial correlation. IWAH = Identification With All Humanity; WVS = World Values Survey.

Table 4. Model fit

				RMSEA			
Model version	*df*	χ^2	*p*[a]	Estimate[b]	90% CI	SRMR	CFI
Model A							
Baseline version	18	137.601	.000	.070**	[.059, .081]	.019	.994
Modification	16	73.862	.000	.051	[.040, .064]	.016	.997
Model B							
Modification	21	108.661	.000	.055	[.045, .066]	.016	.996

Note. df = degrees of freedom; χ^2 = Chi-square test statistic; RMSEA = Root Mean Square Error of Approximation; CI = confidence interval; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index. [a]H_0: $\Sigma = \Sigma(\theta)$. [b]H_0: RMSEA $\leq$ .05. ***p* < .01.

with the first specific factor were negative and comparably weaker (−.298 to −.025).

In Model B, the standardized loadings were also very high for the *G*-factor (.546 to .782) and comparably lower for the specific factors (.093 to .553). Each factor loading was statistically significantly different from zero except the loadings of Item 5 on the second and third specific factors. Moreover, the consistency coefficients of the non-reference indicators with respect to the true variance were between .516 and .806. The item reliabilities were rather high for the single-item measures (.578 to .765). However, the reliability of the RI (.612), the WVS item, and the consistency coefficients of the non-reference indicators (except Item 7) were lower than those in Model A. The correlations of the first (.472) and the third (.775) specific factors with the second specific factor were very large; the correlation between the other two factors was comparably low (.126).

To offer a profound analysis of the modified Models A and B, Table 7 also presents the thresholds of the response options of the individual items.

Discussion

The present study was aimed at determining whether Item 5 from the IWAH scale and the WVS item are appropriate single-item measures of IWAH that can be used when short scales are needed. In general, single items have lower reliabilities than scales with multiple items. This lower reliability can result in lower correlations with criterion variables. When considering these items as single categorical item measures, both items showed comparably high reliability coefficients (.819 and .612). These reliabilities were estimated based on models of confirmatory factor analysis. In confirmatory factor analysis, an item's specific part

Table 5. Parameter estimates and variance coefficients of modified Model A (RI: Item 5)

	G-factor loading			*S*-factor loading		
Item	Estimate	*SE*	Standardized	Estimate	*SE*	Standardized
5	1.000	–	.905	–	–	–
S_1						
1	0.889***	.021	.805	1.000	–	.238
2	0.819***	.025	.741	1.842***	.278	.438
3	0.894***	.020	.809	0.937***	.130	.223
4	0.749***	.024	.677	1.229***	.194	.292
S_2						
6	0.815***	.020	.737	1.000	–	.288
9	0.825***	.023	.747	1.957***	.277	.563
S_3						
7	0.809***	.022	.732	1.000	–	.403
8	0.771***	.025	.697	1.199***	.096	.484
		Consistency		Specificity		
	Rel (Y_i)	Con (Y_i)	Con (τ_i)	Spe (Y_i)	Spe (τ_i)	
1	.704	.648	.920	.057	.081	
2	.741	.549	.741	.192	.259	
3	.704	.654	.929	.050	.071	
4	.544	.458	.842	.085	.156	
5	.819	.819	1	0	0	
6	.626	.543	.867	.083	.133	
7	.699	.536	.767	.162	.232	
8	.720	.486	.675	.234	.325	
9	.875	.558	.638	.317	.362	

Note. G-factor = general factor; *S*-factor = specific factor; *SE* = standard error; Rel = reliability; Y_i = observed variable; τ_i = true score variable; Con = consistency coefficient; Spe = specificity coefficient. Inconsistencies are due to rounding errors. ***p < .001.

might not only comprise measurement error but also true item-specific effects. Therefore, the reliability estimates are lower bounds of reliability.

The correlations between the criterion variables and the single items are very similar to the correlations between the criterion variables and the total scale. This shows that conclusions about the correlations with other variables do not seem to be affected by lower reliabilities. Furthermore, high percentages of the true variance of each non-reference indicator could be explained by a *G*-factor defined by either the WVS item (51.6–80.6%) or Item 5 (63.8–92.9%). Along with the observation that their test-criterion correlations were quite similar to the correlation of their multi-item counterpart, both items can be considered to represent the entire scale sufficiently well. More specifically, for each non-reference indicator besides Item 7, the consistency coefficients were larger for Item 5 as the RI than for the WVS item. Therefore, Item 5 might be more suitable for representing the entire IWAH scale. Nevertheless, the high convergent validity of the WVS item with the items from the IWAH scale indicates that the two instruments measure strongly related constructs. Whereas other studies have argued for omitting Item 5 because of its high loadings on two different IWAH facets, our results indicate that it is this property that might make Item 5 a good representative of the total scale. Its representativeness is also indicated by the comparably small standardized loadings of Item 5 on the specific factors in Model B (.352, .093, .121). Additionally, these loadings show that Item 5 is more closely linked to the first IWAH facet (compared with the other ones).

In contrast to previous studies, our evaluations of fit demonstrated that model versions including three specific factors fit the data best. This finding indicates that it might be worthwhile to distinguish between a dimension of caring and a dimension of moral values in addition to a self-categorization dimension.

As the specificity coefficients (which were larger than 0) indicate that each RI does not explain all interindividual differences in IWAH, a more complete representation of the IWAH scale can be created by adding the item with the highest specificity coefficient for each of the three specific facets to the RI. Thus, if there is a need to capture the multidimensionality of IWAH or to maximize content validity efficiently, a shortened scale consisting of four items (Items 2, 5, 8, 9) would be appropriate. As previous studies have shown that the IWAH facets might be differentially related to other constructs (McFarland & Hornsby, 2015; Reysen & Hackett, 2016), this 4-item scale might be a

Table 6. Parameter estimates and variance coefficients of modified Model B (RI: WVS Item)

	G-factor loading			*S*-factor loading		
Item	Estimate	*SE*	Standardized	Estimate	*SE*	Standardized
WVS	1.000	–	.782	–	–	–
S_1						
1	.939***	.058	.734	1.000	–	.422
2	.800***	.059	.626	1.292***	.105	.545
3	.946***	.057	.740	.977***	.071	.412
4	.698***	.059	.546	1.255***	.114	.529
5	.951***	.060	.743	.834***	.105	.352
S_2						
6	.827***	.055	.647	1.000	–	.495
5	.951***	.060	.743	.189	.158	.093
9	.866***	.058	.677	1.117***	.091	.553
S_3						
7	.973***	.053	.761	1.000	–	.372
5	.951***	.060	.743	.325	.204	.121
8	.863***	.060	.675	1.387***	.115	.516
		Consistency		Specificity		
	Rel (Y_i)	Con (Y_i)	Con (τ_i)	Spe (Y_i)	Spe (τ_i)	
WVS	.612	.612	1	0	0	
1	.717	.539	.752	.178	.248	
2	.689	.392	.569	.297	.431	
3	.717	.548	.764	.170	.237	
4	.578	.298	.516	.280	.484	
5	.759	.552	.727	.206	.271	
6	.664	.419	.631	.245	.369	
7	.718	.579	.806	.138	.192	
8	.722	.456	.632	.266	.368	
9	.765	.458	.599	.306	.400	

Note. *G*-factor = general factor; *S*-factor = specific factor; *SE* = standard error; Rel = reliability; Y_i = observed variable; τ_i = true score variable; Con = consistency coefficient; Spe = specificity coefficient. Inconsistencies are due to rounding errors. ****p* < .001.

Table 7. Thresholds of response options in modified Models A/B

Item	Threshold 1	Threshold 2	Threshold 3	Threshold 4
1	−1.29	−0.42	0.69	1.71
2	−0.57	0.12	0.99	1.78
3	−1.12	−0.31	1.02	1.98
4	−0.03	0.77	1.61	2.13
5	−0.87	−0.09	0.97	1.83
6	−1.30	−0.64	0.30	1.24
7	−1.21	−0.68	0.07	0.97
8	−1.37	−0.80	−0.14	0.67
9	−1.28	−0.69	0.24	1.05
WVS	−1.09	−0.24	1.06	–

Note. WVS = World Values Survey.

useful, efficient tool for analyzing such different associations. If individual assessment of IWAH is intended, presenting the whole scale might go along with a higher precision of estimating personal scores. However, for research purposes, the single items and short scales presented might be an economic way to assess the core of IWAH very efficiently.

Limitations

First, the data used in this study came from only a single country. From the perspective of cross-cultural psychology, the results do not allow conclusions to be drawn about the factor structure of the IWAH scale and efficient ways of measuring IWAH in other countries – especially if they are shaped by other cultures, religions, or political systems (Hamer et al., 2021). Additionally, due to the cross-sectional design, we cannot say whether the same items are appropriate for measuring IWAH and its subcomponents efficiently across different periods of time (McFarland, 2015). Furthermore, the data were collected during the COVID-19 pandemic when global dependences were quite salient, which could have influenced participants' global identification.

Consequently, we recommend that longitudinal studies be conducted and that samples characterized by greater cultural diversity be investigated.

Second, only two items, which were derived theoretically (Item 5) or exploratorily (WVS item), served as the RIs in the analyses in the present study. Descriptive statistical analyses showed that Item 5 was indeed the only item on the scale whose polychoric correlations with all other IWAH items were larger than .60, but the polychoric correlations between the individual items were generally high. Consequently, we cannot determine whether or not another item from the IWAH scale would be more appropriate for defining the *G*-factor if another data set were used. Therefore, replication studies are recommended.

In summary, for future research on GHIC, we recommend replacing the IWAH scale with Item 5 if a single-item measure is needed and with the proposed 4-item solution if a more complex picture is desired. Moreover, the WVS item also represents the IWAH scale well.

Therefore, research questions related to GHIC can also be analyzed by using the WVS data.

References

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science: A call to action. *European Journal of Psychological Assessment, 38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health, 27*(3), 281–291. https://doi.org/10.1093/pubmed/fdi031

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence, 6*(3), Article 42. https://doi.org/10.3390/jintelligence6030042

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Hagel, M. L., Trutzenberg, F., & Eid, M. (2022). Perceived parenting and identification with all humanity: Insights from England and Germany. *Frontiers in Psychology, 13*, Article 924562. https://doi.org/10.3389/fpsyg.2022.924562

Hamer, K., Penczek, M., McFarland, S., Wlodarczyk, A., Łużniak-Piecha, M., Golińska, A., Cadena, L. M., Ibarra, M., Bertin, P., & Delouvée, S. (2021). Identification with all humanity – A test of the factorial structure and measurement invariance of the scale in five countries. *International Journal of Psychology, 56*(1), 157–174. https://doi.org/10.1002/ijop.12678

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., & Puranen, B. (2018). *World Values Survey: Round six-country-pooled datafile version*. JD Systems Institute. https://doi.org/10.14281/18241.8

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology, 95*(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

McFarland, S. (2015). Culture, individual differences, and support for human rights: A general review. *Peace and Conflict: Journal of Peace Psychology, 21*(1), 10–27. https://doi.org/10.1037/pac0000083

McFarland, S., Hackett, J., Hamer, K., Katzarska-Miller, I., Malsch, A., Reese, G., & Reysen, S. (2019). Global human identification and citizenship: A review of psychological studies. *Political Psychology, 40*(1), 141–171. https://doi.org/10.1111/pops.12572

McFarland, S., & Hornsby, W. (2015). An analysis of five measures of global human identification. *European Journal of Social Psychology, 45*(7), 806–817. https://doi.org/10.1002/ejsp.2161

McFarland, S., Webb, M., & Brown, D. (2012). All humanity is my ingroup: A measure and studies of identification with all humanity. *Journal of Personality and Social Psychology, 103*(5), 830–853. https://doi.org/10.1037/a0028724

Meunier, J.-C., & Roskam, I. (2007). Psychometric properties of a Parental Childrearing Behavior Scale for French-speaking parents, children, and adolescents. *European Journal of Psychological Assessment, 23*(2), 113–124. https://doi.org/10.1027/1015-5759.23.2.113

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait-multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology, 59*(1), 195–213. https://doi.org/10.1348/000711005X67490

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? *Journal of Individual Differences, 35*(4), 212–220. https://doi.org/10.1027/1614-0001/a000141

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2014). *Big Five Inventory (BFI-10)*. ZIS – GESIS Leibniz Institute for the Social Sciences.

Reese, G., Proch, J., & Finn, C. (2015). Identification with all humanity: The role of self-definition and self-investment. *European Journal of Social Psychology, 45*(4), 426–440. https://doi.org/10.1002/ejsp.2102

Reitzle, M., Winkler Metzke, C., & Steinhausen, H.-C. (2001). Eltern und Kinder: Der Zürcher Kurzfragebogen zum Erziehungsverhalten (ZKE) [Parents and children: The Zurich Brief Questionnaire for the Assessment of Parental Behaviors]. *Diagnostica, 47*(4), 196–207. https://doi.org/10.1026/0012-1924.47.4.196

Reysen, S., & Hackett, J. (2016). Further examination of the factor structure and validity of the Identification With All Humanity Scale. *Current Psychology, 35*(4), 711–719. https://doi.org/10.1007/s12144-015-9341-y

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rumpold, G., Doering, S., Höfer, S., & Schüssler, G. (2002). Der Fragebogen dysfunktionaler elterlicher Beziehungsstile (FDEB) [(A German version of) the Measure of Parental Style (MOPS)]. *Zeitschrift fur Psychosomatische Medizin und Psychotherapie, 48*(1), 59–74. https://doi.org/10.13109/zptm.2002.48.1.59

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

Su, R., Tay, L., & Diener, E. (2014). The development and validation of the Comprehensive Inventory of Thriving (CIT) and the Brief Inventory of Thriving (BIT). *Applied Psychology: Health and Well-Being, 6*(3), 251–279. https://doi.org/10.1111/aphw.12027

Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science, 4*(1), 1–17. https://doi.org/10.1177/2515245920918253

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*(4), 185–189. https://doi.org/10.1027/1614-0001/a000148

History
Received January 28, 2023
Revision received August 24, 2023
Accepted August 31, 2023
Published online November 6, 2023
EJPA Section / Category Differential Psychology

Publication Ethics
All study participants gave their informed consent before starting to fill out the questionnaire. Local law and the institution's ethics committee did not require ethical approval for the study.

Authorship
Philipp Geisen, statistical analyses, writing – original draft; Friedemann Trutzenberg, data collection, development – research idea, review and editing – original draft, co-supervision of analysis. Michael Eid, supervision (lead role), conceptualization (equal role), funding acquisition, methodology, project administration, review and editing – original draft.

Open Science
We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact *p* values, effect sizes, and 95% confidence or credible intervals.
Open Data: The information needed to reproduce all of the reported results is not openly accessible. The data is available on request from the author(s).
Open Materials: The information needed to reproduce all of the reported methodology is not openly accessible. The material is available on request from the author(s).
Preregistration of Studies and Analysis Plans: This study was not preregistered with an analysis plan.

Funding
Open access publication enabled by Freie Universität Berlin.

ORCID
Michael Eid
https://orcid.org/0000-0001-8920-1412
Philipp Geisen
https://orcid.org/0000-0001-7831-4104
Friedemann Trutzenberg
https://orcid.org/0000-0003-0460-2300

Michael Eid
Department of Education and Psychology
Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin
Germany
eid@zedat.fu-berlin.de

Appendix

Identification With All Humanity Scale

1. How close do you feel to each of the following groups?
   A. Not at all close
   B. Not very close
   C. Just a little or somewhat close
   D. Pretty close
   E. Very close

   People in my community
   Germans
   Europeans
   (Adherents of my religion/denomination)
   People all over the world

2. How often do you use the word "we" to refer to the following groups of people?
   A. Almost never
   B. Rarely
   C. Occasionally
   D. Often
   E. Very often

   People in my community
   Germans
   Europeans
   (Adherents of my religion/denomination)
   People all over the world

3. How much would you say you have in common with the following groups?
   A. Almost nothing in common
   B. Little in common
   C. Some in common
   D. Quite a bit in common
   E. Very much in common

   People in my community
   Germans
   Europeans
   (Adherents of my religion/denomination)
   People all over the world

Please answer all remaining questions using the following choices:
A. Not at all
B. Just a little
C. Somewhat
D. Quite a bit
E. Very much

4. Sometimes people think of those who are not part of their immediate family as "family". To what degree do you think of the following groups of people as "family"?
People in my community
Germans
Europeans
(Adherents of my religion/denomination)
All humans everywhere

5. How much do you identify with (that is, feel a part of, feel love toward, have concern for) each of the following?
People in my community
Germans
Europeans
(Adherents of my religion/denomination)
All humans everywhere

6. How much would you say you care (feel upset, want to help) when bad things happen to:
People in my community
Germans
Europeans
(Adherents of my religion/denomination)
People anywhere in the world

7. How much do you want to be:
A responsible citizen of my community?
A responsible German citizen?
A responsible European citizen?
(a responsible member of my religious community/denomination?)
A responsible citizen of the world?

8. How much do you believe in:
Being loyal to my community?
Being loyal to Germany?
Being loyal to Europe?
(being loyal to my religion/denomination?)
Being loyal to all mankind?

9. When they are in need, how much do you want to help:
People in my community
Germans
Europeans
(Adherents of my religion/denomination)
People all over the world

From "All humanity is my ingroup: A measure and studies of identification with all humanity" by S. McFarland, M. Webb, & D. Brown (2012). *Journal of Personality and Social Psychology*, *103*(5), 830–853. (https://doi.org/10.1037/a0028724). The original scale has been adapted to the groups of Germans, Europeans, and adherents of my religion/denomination.

Reprinted with permission. © American Psychological Association