



Development and validation of three brief versions of the humor styles questionnaire[☆]

Tobias Altmann

Department of Psychology, University of Duisburg-Essen, Universitätsstr. 2, 45141 Essen, Germany

ARTICLE INFO

Keywords:

Humor styles questionnaire
Validation
Brief version
Personality
Well-being

ABSTRACT

The Humor Styles Questionnaire (HSQ) is a frequently applied measure of individual differences in humor usage. However, previous research found several of its items to be unreliable and to provide insufficient information. Its 32-item length may also hinder its application in contexts in which short measures are required to maintain participant motivation, for instance large-scale assessments. In the present paper, we present three brief versions with five, four, and three items per subscale respectively, avoiding the questionable items. The results showed that the factor structure, internal consistency, and test-retest reliability of the original HSQ and the developed brief versions were as expected. Relations between the brief versions and psychological constructs (e.g., self-esteem, well-being, aggression, Big Five) were similar to the original HSQ. The findings support the psychometrical soundness and validity of three brief versions of the HSQ comparable to the original long version. General challenges in the assessment of humor are discussed.

1. Introduction

Humor is an important aspect of social interactions (Gervais & Wilson, 2005). The habitual usage of humor in everyday life has been shown to be related to a number of beneficial outcomes. To name just a few, humor was found to relate positively to mental health (Schneider et al., 2018), subjective well-being (Jiang et al., 2020), and relationship satisfaction (Hall, 2017). Although evidently a pivotal trait, the conceptualization of humor is complex.

1.1. Conceptualization of humor

Conceptualizations of humor can be categorized in three areas: humor production, humor reception, and humor usage. First, humor production describes the ability of a person to create a content that elicits a joyful response in others (Nusbaum et al., 2017). The common method of assessment is to present participants with cartoons but without their humorous captions. The participants are instructed to create humorous captions for these cartoons within a given timeframe. These captions are then judged by others with regard to how funny they are. Although a plausible approach, this method strongly focusses on the ability aspect of humor and therefore confounds with creativity, verbal intelligence, and spontaneity (Kellner & Benedek, 2017). Studies

applying this method usually find that participants find this task to be difficult and most responses receive low funniness rating by the judges (Greengross & Miller, 2011).

Second, humor reception (conceptually similar to humor appreciation) describes the disposition of a person to find joy in potentially humorous situations and contents (El Refaie, 2011). Participants are presented with various types of humorous materials and are instructed to rate how funny this material is to them personally. A major difficulty of this method lies in the breadth of the contents people may find funny. To create a collection of humorous contents that captures all areas of personal preferences is posing a fundamental challenge (Eysenck, 1942).

Third, humor usage describes the everyday application of humor with regard to certain functions of humor. Martin et al. (2003) suggested to differentiate humor usage into four types: affiliative, self-enhancing, aggressive, and self-defeating.

The affiliative humor style is non-threatening and sociable. It fulfills the function of enhancing the relationship by laughing together. An example of this type would be a person who likes to tell jokes. The self-enhancing humor style fulfills the function of enhancing the self in challenging or stressful situations. An example would be a person who tries to take setbacks lightly by finding humorous aspects in his misfortune. These two types have been categorized as adaptive humor styles. They relate positively to extraversion, openness, self-esteem, and

[☆] I have no conflicts of interest to disclose.

E-mail address: tobias.altmann@uni-due.de.

well-being, and negatively to neuroticism (Greengross et al., 2012; Martin et al., 2003; Mendiburo-Seguel et al., 2015; Plessen et al., 2020).

The aggressive humor style aims at enhancing the self at the costs of others. This humor style can be derogative and spiteful with another person as the object of the humor to be laughed at. The self-defeating humor style fulfills the function of enhancing the relationship at the costs of the self. The person tends to put him/herself down to make others laugh at them. These two types have been categorized as maladaptive humor styles. They relate positively to aggression and neuroticism, and negatively to agreeableness and conscientiousness (Greengross et al., 2012; Martin et al., 2003; Mendiburo-Seguel et al., 2015; Plessen et al., 2020).

1.2. The humor styles questionnaire

These four humor styles described above can be assessed using the Humor Styles Questionnaire presented by Martin et al. (2003). This inventory entails eight items for each of the four subscales (i.e., humor styles as described above) and is one of the most widely used inventories of humor styles (Ruch & Heintz, 2016). Its psychometric properties have been validated in different languages and generally support its psychometric soundness (Ruch & Heintz, 2016; Schermer et al., 2019; Sirigatti et al., 2014; Torres-Marín et al., 2018).

However, a detailed analysis of the 32 items using item response theory analyses (Silvia & Rodriguez, 2020) revealed that several of its items showed substantial flaws. The most important shortcomings seem to be that several items showed poor discrimination (insufficient item information) and high local dependence (correlated errors between items). Furthermore, with eight items per subscale and 32 items in total, the HSQ may be too long for certain assessment requirements. This may be the case in contexts in which short measures are required to maintain participant motivation such as in large-scale assessments or panel studies. Here, most constructs are assessed using short inventories or brief versions of longer measures, often with three or even less items per construct (Gosling et al., 2003; Rammstedt & John, 2007). Thus, it may be argued that the HSQ might benefit from a reduction of items (Silvia & Rodriguez, 2020) and its eight items per subscale allow for such a reduction (Scheel et al., 2016).

Therefore, the goal of the present study is to develop and evaluate a set of brief versions of the HSQ. In Study 1, we will create three brief versions with five, four, and three items per subscale respectively and test their factorial structure and relations to age and gender. In Study 2, we test the construct validity of these brief versions regarding other personality constructs and their test-retest reliability. We report how we determined our sample size, all data exclusions, all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, all analyses including all tested models, *p*-values, and effect sizes.

2. Study 1

The goal of Study 1 was to create a set of brief versions of the HSQ by selecting items which are psychometrically sound and most indicative of the latent construct (i.e., humor style). We expected to find at least five items per subscale with good psychometric properties. We also expected to be able to create three brief versions (with five, four, and three items per subscale) which achieve good fit of the structural models in confirmatory factor analysis.

Lastly, we expected to find similar gender differences and associations with age in the original HSQ and the brief versions. Typically, men score higher on all four subscales (Martin et al., 2003; Ruch & Heintz, 2016), therefore we expected to find the same in the present sample in both the original HSQ and the brief versions alike. With regard to the association with age, Martin et al. (2003) showed that younger participants scored higher on the affiliative and the aggressive humor styles, therefore, we expected these two negative associations in the original

and the brief versions alike.

2.1. Methods

To develop the brief versions, we used the open dataset provided by Silvia and Rodriguez (2020) which entails 1210 participants with HSQ data using a five-point Likert scale. However, we excluded participants who provided incomplete data prior to first data analysis ($n = 73$). Thus, the final sample included 1137 participants, 46.0 % women, mean age 30.9 years ($SD = 12.0$). Details about the dataset can be found in the respective publication. In addition, we also refer to the item response theory-based analyses provided by Silvia and Rodriguez (2020) in our first step of item selection as detailed below.

2.2. Results

2.2.1. Preselection of items

To create a selection of psychometrically fit items or to omit unfit items, we applied the analyses presented by Silvia and Rodriguez (2020) regarding the quality of the individual HSQ items. To do so, we used three criteria: discrimination, local dependence, and disorderedness. First, discrimination describes the level of information an item contains and thereby its ability to separate trait-wise similar participants. We used the items' discrimination values to rank order the items of each of the four HSQ subscales separately. Second, local dependence describes unwanted residual correlations between two items. Of each item showing local dependence with another (specifically: Items 4 with 24, 5 with 29, 6 with 30, 13 with 25, 24 with 32), only one of the two would be included in the brief versions. Third, an item is marked as disordered if the item's Likert scale steps do not adequately represent the participants' ability levels (as a simplified example: participants with a trait level of "4" frequently use the item rating "3" whereas participants with a trait level of "3" frequently use the rating "4"). These items (specifically: Items 11, 19, 25, 27, and 28) were omitted from the list of potential items for the brief versions due to this criterion.

To provide an example of the application of these criteria in the item preselection process, we briefly explicate the application on the affiliative humor subscale. Item 25 ("I don't often joke around with my friends") showed the best discrimination score of 1.852 compared with the other items of this subscale. Therefore, it was ranked first. However, this item also showed signs of disorderedness and local dependence with item 13 and was therefore discarded as psychometrically unacceptable. Item 2 ("If I am feeling depressed, I can usually cheer myself up with humor") ranked second regarding the discrimination scores and did not show any issues with disorderedness or local dependence. It was therefore retained.

This process was exercised with all items in all subscales. The resulting number of preselected items per HSQ subscale were six for affiliative, seven for self-enhancing, five for aggressive, and six for self-defeating. Table SM1 in the Supplementary Material provides details of the scores and criteria which led to the preselection described.

2.2.2. Development of the three brief versions

We decided to use the smallest number (i.e., five) as the number of items for each subscale to create the first brief version of the HSQ. We selected the best five items (for those subscales with more than five items preselected) using the discrimination values rank order. These five items are listed in Table 1 and constitute the first brief version with five items per subscale as well as the basis for the development of the four-item and three-item per subscale versions.

We further reduced the number of items to also form a four-item and a three-item brief version of the HSQ. Three items were chosen as the minimum to enable other researchers to conduct a one-factor confirmatory factor analysis with a saturated model. To create these versions, we cut the sample ($N = 1137$) in half based on a newly created dichotomous random variable ($N_1 = 582$, $N_2 = 555$). With the first

Table 1
Five highest loading items in confirmatory factor analysis based on the complete HSQ.

Affiliative		Self-enhancing		Aggressive		Self-Defeating	
Nr	Loading	Nr	Loading	Nr	Loading	Nr	Loading
01	0.72	02	0.76	03	0.78	04	0.70
05	0.73	10	0.74	07	0.48	08	0.89
13	0.61	14	0.91	15	0.69	12	0.73
17	0.76	18	0.66	23	0.53	20	0.77
21	0.52	26	0.72	31	0.71	32	0.83

sample half, we calculated a confirmatory factor analysis including all items of the HSQ using the R package “lavaan” (Rosseel, 2012) and applying the weighted least square mean and variance adjusted estimator due to the ordinal data structure (Brauer et al., 2023; Li, 2016). The model fitted the data well ($\chi^2(458, N = 582) = 1147.24, p < .001$, RMSEA = 0.056, CFI = 0.927, SRMR = 0.065; robust estimation) which is in line with previous research (Ruch & Heintz, 2016). The resulting loadings (presented in Table 1) indicate how well each of the selected five item per subscale represents the respective latent subscale as conceptualized in the original version of the HSQ with eight items per subscale.

We then ranked these five preselected items based on their loadings to select the best four items for the four-item version and the best three items for the three-item version. To evaluate the psychometric soundness of the three brief versions created with the first sample half, we used the second sample half ($N_2 = 555$) to calculate confirmatory factor analyses for each brief version (see Table 2). All models fitted the data better than the original HSQ version. Table 3 displays the descriptive statistics and internal consistency of the original eight-item version and the three brief versions for each of the four humor styles (i.e., HSQ subscales) including the final selection items per brief version. Table SM2 in the Supplementary Material includes a more detailed list of these items.

We then tested for gender differences and associations with age. As expected, men scored significantly higher on all subscales and in all questionnaire versions with only two exceptions. The gender difference in self-enhancing humor was just not significant in the four-item version ($p = .0583$) and the gender difference in affiliative humor was just not significant in the three-item version ($p = .0576$). With regard to age, we found negative associations with the affiliative and the aggressive humor styles in all versions as expected (Table 4).

2.3. Discussion

Supported by previous work (Silvia & Rodriguez, 2020), we were able to create three brief versions of the HSQ and thereby omit its psychometrically questionable items. The brief versions showed good psychometric properties which were comparable with the properties of the original version. Of course, a certain degree of loss in information occurred parallel to the reduction of items from eight to five to four to three. The distributions of mean scores in the subscales were somewhat wider as indicated by the larger standard deviation. Similarly, internal consistency decreased slightly when comparing the original with the brief versions. This decrease was largest for the affiliative humor style (original: 0.85, three-item versions: 0.72) and smallest for the self-

Table 2
Confirmatory factor analyses of the original and the brief versions.

Brief version	χ^2	df	p	RMSEA	CFI	SRMR
8-item original	1147.24	458	<.001	0.056	0.927	0.065
5-item version	338.70	164	<.001	0.042	0.981	0.051
4-item version	220.11	98	<.001	0.042	0.970	0.048
3-item version	111.00	48	<.001	0.041	0.987	0.043

Note. Fit measures based on robust estimation.

Table 3
Descriptive statistics and coefficients of internal consistency.

Subscale and version (HSQ item numbers)	M (SD)	Internal consistency	
		Alpha	Omega
Affiliative humor style			
Original 8-item version	3.97 (0.70)	0.85	0.85
5-item version (01,05,13,17,21)	4.05 (0.72)	0.81	0.81
4-item version (01,05,13,17)	3.98 (0.77)	0.77	0.77
3-item version (01,05,17)	3.86 (0.83)	0.72	0.73
Self-enhancing humor style			
Original 8-item version	3.40 (0.72)	0.83	0.83
5-item version (02,10,14,18,26)	3.26 (0.85)	0.82	0.82
4-item version (02,10,14,26)	3.34 (0.85)	0.77	0.77
3-item version (02,10,14)	3.24 (0.90)	0.72	0.72
Aggressive humor style			
Original 8-item version	2.80 (0.76)	0.79	0.79
5-item version (03,07,15,23,31)	2.89 (0.84)	0.74	0.74
4-item version (03,15,23,31)	2.94 (0.90)	0.72	0.72
3-item version (03,15,31)	2.84 (0.96)	0.67	0.68
Self-defeating humor style			
Original 8-item version	2.76 (0.76)	0.82	0.82
5-item version (04,08,12,20,32)	2.71 (0.85)	0.80	0.80
4-item version (08,12,20,32)	2.69 (0.90)	0.78	0.78
3-item version (08,20,32)	2.59 (0.95)	0.77	0.78

Note. Numbers in parentheses denote the items numbers of the HSQ selected to form the respective brief version.

Table 4
Correlations between HSQ subscales and age for the original and the brief versions.

Humor style	HSQ version			
	Original	5-item version	4-item version	3-item version
Affiliative	-0.14**	-0.13**	-0.14**	-0.09*
Self-enhancing	-0.03	-0.02	-0.01	-0.03
Aggressive	-0.21**	-0.23**	-0.24**	-0.24**
Self-defeating	-0.06	-0.05	-0.03	-0.03

* $p < .05$.

** $p < .01$.

defeating humor style (original: 0.82, three-item version: 0.77). In all cases, we evaluate the drop in internal consistency as adequate given the reduction of item number (Nunnally, 1967). In contrast, structural models in confirmatory factor analysis increased in fit. This is also to be expected because we selected the best indicators of the latent construct but with a smaller number of items, thus reducing the degrees of freedom. Gender differences and associations with age also resulted in highly comparable results across all HSQ versions.

Although these first results look promising, they are limited to the psychometric properties inherent in the items and their subscales (except for gender and age). A second study is needed to evaluate the validity of the brief version with regard to their ability to explain variance in other relevant psychological constructs as good as the original HSQ version to be deemed parallel brief versions.

3. Study 2

The goal of Study 2 was to evaluate the validity and to further evaluate the reliability of the three brief versions developed in Study 1. To achieve this, we tested the relations of the original and the three brief versions with other psychological constructs (e.g., Big Five, self-esteem, aggression, subjective well-being) as in previous studies (Martin et al., 2003; Sirigatti et al., 2014) and assessed their test-retest reliability over a two-week span.

With regard to validity, we expected the two adaptive humor styles (i.e., affiliative and self-enhancing humor styles) to relate positively to the following socially desirable constructs: affiliation motive, agreeableness, conscientiousness, extraversion, self-esteem, psychological

well-being, optimism, and positive affect. Conversely, we expected negative associations between the two adaptive humor styles and the following socially undesirable outcomes: neuroticism, aggression, pessimism, and negative affect.

The opposite pattern of results was expected for the two maladaptive humor styles (i.e., aggressive and self-defeating humor styles). Specifically, we expected the two maladaptive styles to relate negatively to the aforementioned socially desirable constructs (affiliation motive, agreeableness, conscientiousness, extraversion, self-esteem, psychological well-being, optimism, and positive affect). Parallel to above, we expected positive associations between the two maladaptive humor styles and the aforementioned socially undesirable outcomes (neuroticism, aggression, pessimism, and negative affect).

These associations have been found in previous studies as described above (Greengross et al., 2012; Martin et al., 2003; Mendiburo-Seguel et al., 2015; Plessen et al., 2020). However, these previously reported results are partially inconsistent and varying in the strength of these effects. We therefore refrained from specifying statistical hypotheses. Instead, we expected the brief versions created in Study 1 to correlate with these constructs only if the original HSQ version correlates with them in the present sample and expected similar effect sizes in the present sample across the HSQ versions.

Also, due to the previously found gender differences and associations with age, we control for gender and age in all correlations regarding validity. With regard to reliability, we expected to find sufficient test-retest reliability above 0.50 in all subscales and all versions.

In addition, we need to ensure that the HSQ versions of Study 1 (English-speaking sample) and Study 2 (German-speaking sample) are comparable. Therefore, we tested for measurement invariance between the English and German HSQ versions.

3.1. Methods

3.1.1. Sample and procedure

Participants were recruited at the local campus at the university of Duisburg-Essen, Germany, and received partial course credit for participation. Data were assessed in supervised group sessions with up to eight participants. We calculated the required sample size using GPOWER version 3.1.9.7 (Faul et al., 2009). Assuming at least medium effect sizes for the correlational analyses of 0.30 (Cohen, 1988), alpha of 0.05, and power of 0.80, the required sample size was 67 for one-tailed hypotheses. Therefore, the applied assessment stopping rule was reaching 67 participants after applying the following exclusion criteria: incomplete data, implausible data, less than excellent control of the German language (self-rated). A total of 96 participants took part in the assessment of which 29 had to be excluded, yielding a final sample of 67. Of those, 43 were women. The mean age was 23.1 years ($SD = 7.1$). To assess test-retest reliability, the HSQ was administered a second time with two weeks between assessments (mean time difference between assessments was 14.0 days, $SD = 1.8$).

For the assessment of measurement invariance, 198 additional participants were recruited using the same recruitment and assessment procedure as above but assessing only the HSQ. Combined with the HSQ data from the 67 participants described above, this sample contained 265 participants of which 69 % were women. The mean age was 25.3 years ($SD = 10.2$). The data can be obtained from the author upon request.

3.1.2. Measures

German versions of the measures described below were used as well as a five-point Likert scale (1 = “does not apply to me” to 5 = “does apply to me”).

3.1.2.1. HSQ. The standard HSQ (Martin et al., 2003; Ruch & Heintz, 2016) was applied with a total of 32 items (eight per subscale). Please

note that we use a five-point scale here as well because it tends to outperform the commonly used seven-point scale as shown by Silvia and Rodriguez (2020). Internal consistency in the present sample was comparable with reports in previous works (Martin et al., 2003) as well as with the findings reported in Study 1 above. Cronbach alphas for the original eight-item version as well as for the three brief versions with five, four, and three items per subscale were as follows: 0.83/0.78/0.74/0.67 for the affiliative, 0.79/0.78/0.70/0.73 for the self-enhancing, 0.75/0.74/0.74/0.68 for the aggressive, and 0.86/0.83/0.82/0.81 for the self-defeating humor style.

3.1.2.2. Unified Motive Scale: affiliation. To assess the affiliation motive, we used the Unified Motive Scale by Schönbrodt and Gerstenberg (2012). The affiliation subscale contains ten items (e.g., “I try to be in the company of friends as much as possible.”) which showed high internal consistency 0.93. We also assessed the achievement and power subscales with ten items each but since we did not have any hypotheses related to them, we did not include them in the analyses.

3.1.2.3. Aggression questionnaire. To assess aggression, we used the total score of the Aggression Questionnaire (Buss & Perry, 1992; von Collani & Werner, 2005) with 29 items (e.g., “I sometimes feel like a powder keg ready to explode.”). Internal consistency was good with 0.79.

3.1.2.4. Big Five inventory-short. To assess the Big Five, we used the Big Five Inventory-Short (Schupp & Gerlitz, 2008) containing three items per dimension. Internal consistency was acceptable to good for the five dimensions agreeableness (e.g., “I see myself as someone who has a forgiving nature”): 0.61, conscientiousness (e.g., “I see myself as someone who does a thorough job”): 0.80, extraversion (e.g., “I see myself as someone who is talkative”): 0.84, neuroticism (e.g., “I see myself as someone who worries a lot”): 0.76, and openness (e.g., “I see myself as someone who is original, comes up with new ideas”): 0.70.

3.1.2.5. Life Orientation Test-Revised (LOT-R). To assess optimism and pessimism, we used the ten item Life Orientation Test-Revised (Glaesmer et al., 2008; Scheier et al., 1994) with three items per subscale plus four filler items. Internal consistency was acceptable for optimism (e.g., “I’m always optimistic about my future.”): 0.77, and for pessimism (e.g., “If something can go wrong for me, it will.”): 0.73.

3.1.2.6. Positive and negative affect schedule. To assess situational positive and negative affect, we used the 20-item Positive and Negative Affect Schedule (Breyer & Bluemke, 2016; Watson & Clark, 1988) asking participants how they feel at the moment. Internal consistency was acceptable to good for positive affect (e.g., “strong”) with 0.72, and negative affect (e.g., “distressed”) with 0.84.

3.1.2.7. Psychological Well-Being (PWB). To assess psychological adjustment, we used the Psychological Well-Being Scale (Risch et al., 2005; Ryff & Keyes, 1995) with 18 items (e.g., “I like most parts of my personality.”). The total score showed sufficient internal consistency with 0.78.

3.1.2.8. Rosenberg Self-Esteem Scale (RSES). To assess self-esteem, we used the Rosenberg Self-Esteem Scale (Ferring & Filipp, 1996; Rosenberg, 1965) with 10 items (e.g., “I take a positive attitude toward myself.”). The total score reached a good level of internal consistency with 0.84.

3.2. Results

3.2.1. Validity

Validity of the three brief versions (five, four, and three items per

subscale respectively) was evaluated based on their correlations with the other psychological constructs. The correlations (controlling for age and gender) are summarized in Table 5.1 for the adaptive humor styles (affiliative and self-enhancing) and Table 5.2 for the maladaptive humor styles (aggressive and self-defeating).

Across all HSQ versions and subscales, the relations between the original HSQ version and the three brief versions were generally similar regarding their effect sizes (please note that we did not compare the correlations for significant differences because the present sample size would require a difference in correlations of around 0.25 to reach significance, and all differences between the HSQ versions were smaller than that). Also, the effect sizes found in the present sample were generally similar to those found in other studies (Greengross et al., 2012; Martin et al., 2003; Mendiburo-Seguel et al., 2015; Plessen et al., 2020). Against our general expectations, the general pattern of positive relations between the adaptive humor styles with desirable constructs and negative relations with undesirable constructs (and vice versa for the maladaptive humor styles) was not found in each case. However, when the original HSQ version correlated with a construct in the present sample, the brief versions did so too in almost all cases.

The only three exceptions to this pattern are the relations between the affiliative humor style with self-esteem and with pessimism, and between the self-defeating humor style and psychological well-being. In all three cases, the five-item and the four-item version correlated significantly as did the original HSQ version, and only the three-item version did not. Also in all three cases, this may be due to the generally small effect size plus the slight decrease in the effect size below the level of detectability of the present sample so that correlations below 0.20 could not be detected (see power analysis above). In reverse, there were five instances in which at least one of the brief versions correlated as expected but the original HSQ version did not: self-enhancing humor style with affiliation and with extraversion, aggressive humor style with psychological well-being and with positive affect, and self-defeating humor style with openness and positive affect. In sum, the brief versions related at least equally well with relevant psychological constructs.

3.2.2. Test-retest reliability

Test-retest reliability was assessed across a two-week span. Table 6 shows the respective correlation coefficients for each subscale across the two measurement occasions. In general, all coefficients were large and changes from the original eight-item measure to the brief versions remained small. The largest difference of 0.13 was found for the aggressive humor style subscale between the eight-item and the four-

Table 5.1

Correlations between the two adaptive hsq subscales (affiliative and self-enhancing) and related constructs for each HSQ version.

	Affiliative humor style				Self-enhancing humor style			
	8-i	5-i	4-i	3-i	8-i	5-i	4-i	3-i
Affiliation	0.36**	0.42**	0.31**	0.23*	0.16	0.22*	0.19	0.24*
Aggression	0.01	0.00	0.01	0.07	-0.12	-0.05	-0.05	-0.05
Big Five								
Agreeableness	-0.03	-0.03	-0.09	-0.13	0.07	0.07	0.07	0.08
Conscientiousness	-0.11	-0.10	-0.09	-0.11	0.13	0.07	0.08	0.10
Extraversion	0.39**	0.42**	0.37**	0.31**	0.14	0.14	0.16	0.21*
Neuroticism	-0.16	-0.16	-0.18	-0.18	-0.18	-0.12	-0.15	-0.20
Openness	0.35**	0.35**	0.33**	0.30**	0.09	0.01	0.04	0.06
Self-esteem	0.24*	0.24*	0.21*	0.20	0.45**	0.42**	0.45**	0.43**
Psych. well-being	0.41**	0.37**	0.34**	0.30**	0.44**	0.40**	0.44**	0.44**
Life Orientation								
Optimism	0.13	0.19	0.12	0.08	0.28*	0.34**	0.33**	0.34**
Pessimism	-0.28*	-0.24*	-0.22*	-0.18	-0.31**	-0.22*	-0.25*	-0.26*
Affect								
Negative affect	0.12	0.15	0.14	0.10	0.18	0.18	0.16	0.19
Positive affect	-0.47**	-0.45**	-0.47**	-0.42**	-0.30**	-0.22*	-0.22*	-0.24*

Note. 8-i denotes the (original) 8-item per subscale version of the HSQ, 5-i denotes the brief 5-item per subscale version, and so forth.

* p < .05.
 ** p < .01.

item version (see Table 6). Table 7 reports the means and standard deviations of all HSQ versions.

3.2.3. Measurement invariance

To ensure that the HSQ brief versions created in Study 1 using an English-speaking sample are comparable with the same brief versions validated in Study 2 using a German-speaking sample, we tested for measurement invariance between the English HSQ (data as used in Study 1) and German HSQ (data as described above for the assessment of measurement invariance). We used the five-item version in both samples as the most relevant brief version to be tested as it also contains the items of the remaining brief versions.

For the present purpose, four models were calculated: the configural model and three models consecutively restricting the parameters loadings, residual variances, and latent factor covariances to be equal across the samples (each model including the restrictions of the former). The models were tested using the R package “lavaan” (Rosseel, 2012). The results in Table 8 show that all ΔCFI (calculated between consecutive models) were smaller than 0.01 and all ΔRMSEA were smaller than 0.015 (Hirschfeld & Brachel, 2014), so that measurement invariance was supported.

3.3. Discussion

Higher reduction from eight (original version) to five to four to three items per dimension in the brief versions can be expected to be accompanied by a drop in validity and test-retest reliability due to the increase in measurement error with smaller item number. Such a drop was found regarding reliability in all subscales but always only to a small degree. A drop was also found regarding validity in three instances but also only to a small degree. In reverse, we found five instances in which the brief versions correlated with relevant constructs as expected from previous studies, but the original HSQ version did not (e.g., regarding positive affect and psychological well-being). In sum, we conclude that the brief versions performed well with regard to validity and reliability.

4. General discussion

4.1. Summary of findings

In the present study, we developed a set of three brief versions of the HSQ with five, four, and three items per subscale (i.e., 20, 16, and 12 items in total) respectively. The brief versions performed well (i.e.,

Table 5.2
Correlations between the two maladaptive HSQ subscales (aggressive and self-defeating) and related constructs for each HSQ version.

	Aggressive humor style				Self-defeating humor style			
	8-i	5-i	4-i	3-i	8-i	5-i	4-i	3-i
Affiliation	0.03	-0.05	-0.01	-0.05	-0.05	-0.05	-0.06	-0.06
Aggression	0.18	0.12	0.06	0.07	0.26*	0.27*	0.29**	0.30**
Big Five								
Agreeableness	-0.54**	-0.42**	-0.32**	-0.33**	0.04	0.02	-0.01	0.03
Conscientiousness	-0.37**	-0.37**	-0.41**	-0.34**	-0.25*	-0.25*	-0.25*	-0.21*
Extraversion	-0.03	-0.07	-0.05	-0.09	-0.17	-0.16	-0.16	-0.16
Neuroticism	0.02	0.13	0.09	0.08	0.33**	0.36**	0.37**	0.33**
Openness	0.04	-0.08	-0.11	-0.12	-0.19	-0.19	-0.20	-0.21*
Self-esteem	0.00	-0.10	-0.07	-0.09	-0.32**	-0.28*	-0.27*	-0.24*
Psych. well-being	-0.19	-0.26*	-0.25*	-0.27*	-0.31**	-0.26*	-0.25*	-0.20
Life Orientation								
Optimism	0.05	-0.06	0.02	0.04	-0.23*	-0.22*	-0.25*	-0.23*
Pessimism	0.03	0.10	0.10	0.08	0.35**	0.29**	0.24*	0.22*
Affect								
Negative affect	0.05	-0.02	-0.01	-0.07	-0.18	-0.18	-0.19	-0.16
Positive affect	0.18	0.30**	0.31**	0.32**	0.20	0.21	0.22*	0.20

Note. 8-i denotes the (original) 8-item per subscale version of the HSQ, 5-i denotes the brief 5-item per subscale version, and so forth.

* $p < .05$.

** $p < .01$.

Table 6
Test-retest reliability for each subscale and version.

Humor style subscale	Original	Brief version		
	8-item version	5-item version	4-item version	3-item version
Affiliative	0.78	0.76	0.72	0.67
Self-enhancing	0.85	0.86	0.85	0.83
Aggressive	0.85	0.75	0.72	0.74
Self-defeating	0.88	0.83	0.84	0.82

Note. All $p < .001$.

similar to the original HSQ measure) regarding factorial structure, internal consistency, relations to age and gender, relations to other relevant psychological constructs (validity), and two-week test-retest correlations (reliability). Although the reduction of items from the original eight items per subscale to five, four, and three items respectively was substantial and therefore expected to be accompanied by a drop in scale quality, the brief versions showed only slight reductions in quality for instance regarding internal consistency and test-retest reliability, but also showed increased quality, for instance, regarding factorial structure and validity.

Table 7
Means (and standard deviations in parentheses) of the original and the brief versions.

Humor style subscale	Original		Brief version					
	8-item version		5-item version		4-item Version		3-item version	
Affiliative	4.20	(0.54)	4.32	(0.59)	4.27	(0.62)	4.15	(0.69)
Self-enhancing	3.35	(0.76)	3.22	(0.84)	3.29	(0.82)	3.24	(0.86)
Aggressive	2.45	(0.72)	2.48	(0.78)	2.47	(0.81)	2.40	(0.86)
Self-defeating	2.59	(0.80)	2.53	(0.92)	2.51	(0.92)	2.39	(0.95)

Note. Using the first measurement occasion.

Table 8
Invariance tests between the english and German samples.

Model	χ^2 (df)	SRMR	RMSEA [90 % CI]	CFI	Δ CFI	Decision
Configural	3280.63 (916)	0.066	0.061 [0.058–0.063]	0.836	–	–
+ equal loadings	3381.83 (948)	0.073	0.061 [0.058–0.063]	0.831	0.005	Accept
+ equal residuals	3481.37 (980)	0.074	0.060 [0.058–0.063]	0.826	0.005	Accept
+ equal covariances	3482.98 (986)	0.075	0.060 [0.058–0.062]	0.827	0.001	Accept

4.2. Single-item measurement

Although statistically possible, further reductions to two or even one item per subscale or even to one item to assess humor as such may be argued to be inappropriate (for a recent discussion of this issue, see [Allen et al., 2022](#)). There is a growing amount of research presenting one-item measures of complex psychological constructs such as empathy ([Konrath et al., 2018](#)) and narcissism ([Konrath et al., 2014](#)). However, using such a small number of items might lead to blending independent psychological mechanisms and diminishes or even annuls the possibility of controlling for measurement error. In addition, this is all the more problematic if the construct is wide in scope or even differentiated in fundamentally different aspects (e.g., cognitive versus affective empathy, or grandiose versus vulnerable narcissisms; [Hall & Schwartz, 2022](#); [Wink, 1991](#)), which also applies to the construct of humor usage with its four humor usage styles ([Martin et al., 2003](#)). A single-item measure of humor could be a sentence such as “I am a humorous person” to which the participants have to indicate their personal level of agreement to describe their self-views. However, given the complexity of humor (humor production, humor reception, and four types of humor usage, as introduced above) such a measure would not be recommendable (however, see [Sulejmanov & Renner, 2017](#) for an interesting suggestion of a peer-rated single-item witness measure).

4.3. Challenges in the assessment of humor

In general, the assessment of humor is not without its challenges. As mentioned above, humor can be conceptualized in quite different ways, mainly as the ability to produce humor spontaneously (i.e., humor production; Nusbaum et al., 2017), as the disposition to find joy in different contents or situations (i.e., humor reception or humor appreciation; El Refaie, 2011), or as the disposition or inclination of using humor in everyday life (i.e., humor styles; Martin et al., 2003). However, assessments of humor production are usually confounded with creativity and spontaneity, assessments of humor reception are highly limited compared to the width of humor-related experiences in humans, and assessments of humor styles suffer from all the problems of questionnaire assessments, for instance, self-deception and impression management as two components of socially desirable responding (Paulhus, 1984) may affect the questionnaire self-rating especially of the maladaptive aggressive and self-defeating humor styles (Cann & Matson, 2014). Thus, fundamental challenges lie in the necessity of optimizing each of these standard approaches toward assessing humor.

In light of the present study on brief versions of the HSQ, our brief versions avoid some of the existing psychometric problems of several HSQ items. However, the HSQ – or in more general terms the operationalization of humor using four styles of humor as suggested by Martin et al. (2003) – may be in need of a thorough revision, for instance, with regard to social desirability, psychometric properties, item information, and item phrasing (Ruch & Heintz, 2017; Silvia & Rodriguez, 2020). Previous research showed that several items of the HSQ contain content related to other constructs, for instance, items assessing the affiliative humor style also contain content related to extraversion, rendering correlations between the two constructs trivial (Ruch & Heintz, 2017). The goal of the present study was not to present a solution to these concerns with the validity of the HSQ, but simply to present brief versions of the HSQ by avoiding some of its problematic items. Further studies are needed to optimize or rethink the assessment of humor styles using questionnaires.

4.4. Limitations

Regarding the limitations of the present study, four issues may need to be discussed. First, the present study uses data from English-speaking and German-speaking samples. Although we provide evidence of measurement invariance, cultural differences between the samples may introduce a cultural bias (Schermer et al., 2019) and limit the validity of the present findings (Hofstede, 2011). It may be advised to replicate the present findings in future research.

Second, the generalizability of the findings of Study 2 may be limited due to the present sample. Because the sample was recruited at a university campus, the sample is younger and of higher education than the general population. Replications using samples more diverse in age and educational backgrounds may be needed to support the generalizability of the present findings.

Third, the present study is limited by using questionnaire data (Martin & Ford, 2018). Future studies may benefit from including test measures of humor to complement the questionnaire measures.

Lastly, as mentioned above, the HSQ has been criticized regarding its validity (Ruch & Heintz, 2017; Silvia & Rodriguez, 2020). The validity of the brief versions presented here is limited by the validity of the original HSQ. We tried to avoid the psychometrically problematic items as described above, but future studies may need to develop new approaches of the assessment of humor to fully overcome these limitations.

4.5. Conclusion

In conclusion, the brief versions of the HSQ presented and evaluated in the present study appear to be adequate alternatives to administering the complete HSQ. This may enable the application of this measure in a

wider field of studies, including studies with higher restrictions regarding item numbers such as large-scale assessments.

CRediT authorship contribution statement

Tobias Altmann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

We have no known conflict of interest to declare.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.actpsy.2024.104141>.

References

- Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment*, 38(1), 1–5. <https://doi.org/10.1027/1015-5759/a000699>
- Brauer, K., Ranger, J., & Ziegler, M. (2023). Confirmatory factor analyses in psychological test adaptation and development. *Psychological Test Adaptation and Development*, 4(1), 4–12. <https://doi.org/10.1027/2698-1866/a000034>
- Breyer, B., & Bluemke, M. (2016). Deutsche version der Positive and Negative Affect Schedule PANAS (GESIS panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. <https://doi.org/10.6102/zis242>
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. <https://doi.org/10.1037//0022-3514.63.3.452>
- Cann, A., & Matson, C. (2014). Sense of humor and social desirability: Understanding how humor styles are perceived. *Personality and Individual Differences*, 66, 176–180. <https://doi.org/10.1016/j.paid.2014.03.029>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- El Refaie, E. (2011). The pragmatics of humor reception: Young people's responses to a newspaper cartoon. *Humor - International Journal of Humor Research*, 24(1), 87–108. <https://doi.org/10.1515/humr.2011.005>
- Eysenck, H. J. (1942). The appreciation of humour: An experimental and theoretical study. *British Journal of Psychology. General Section*, 32(4), 295–309. <https://doi.org/10.1111/j.2044-8295.1942.tb01027.x>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Ferring, D., & Philipp, S.-H. (1996). Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala [Measurement of self-esteem: Reliability, validity, and stability of the Rosenberg Scale]. *Diagnostica*, 42(3), 284–292.
- Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology*, 80(4), 395–430. <https://doi.org/10.1086/498281>
- Glaesmer, H., Hoyer, J., Klotsche, J., & Herzberg, P. Y. (2008). Die deutsche Version des Life-Orientiation-Tests (LOT-R) zum dispositionellen Optimismus und Pessimismus. *Zeitschrift für Gesundheitspsychologie*, 16(1), 26–31. <https://doi.org/10.1026/0943-8149.16.1.26>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/s0092-6566\(03\)00046-1](https://doi.org/10.1016/s0092-6566(03)00046-1)
- Greengross, G., Martin, R. A., & Miller, G. (2012). Personality traits, intelligence, humor styles, and humor production ability of professional stand-up comedians compared to college students. *Psychology of Aesthetics, Creativity, and the Arts*, 6(1), 74–82. <https://doi.org/10.1037/a0025774>
- Greengross, G., & Miller, G. (2011). Humor ability reveals intelligence, predicts mating success, and is higher in males. *Intelligence*, 39(4), 188–192. <https://doi.org/10.1016/j.intell.2011.03.006>
- Hall, J. A. (2017). Humor in romantic relationships: A meta-analysis. *Personal Relationships*, 24(2), 306–322. <https://doi.org/10.1111/perc.12183>
- Hall, J. A., & Schwartz, R. (2022). Empathy, an important but problematic concept. *The Journal of Social Psychology*, 162(1), 1–6. <https://doi.org/10.1080/00224545.2021.2004670>
- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, 2(1). <https://doi.org/10.9707/2307-0919.1014>

- Jiang, F., Lu, S., Jiang, T., & Jia, H. (2020). Does the relation between humor styles and subjective well-being vary across culture and age? A meta-analysis. *Frontiers in Psychology, 11*, 2213. <https://doi.org/10.3389/fpsyg.2020.02213>
- Kellner, R., & Benedek, M. (2017). The role of creative potential and intelligence for humor production. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 52–58. <https://doi.org/10.1037/aca0000065>
- Konrath, S., Meier, B. P., & Bushman, B. J. (2014). Development and validation of the Single Item Narcissism Scale (SINS). *PLoS One, 9*(8), Article e103469. <https://doi.org/10.1371/journal.pone.0103469>
- Konrath, S., Meier, B. P., & Bushman, B. J. (2018). Development and validation of the Single Item Trait Empathy Scale (SITES). *Journal of Research in Personality, 73*, 111–122. <https://doi.org/10.1016/j.jrp.2017.11.009>
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Martin, R. A., & Ford, T. E. (2018). The personality psychology of humor. In *The psychology of humor* (pp. 99–140). <https://doi.org/10.1016/b978-0-12-812143-6.00004-7>
- Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality, 37*(1), 48–75. [https://doi.org/10.1016/s0092-6566\(02\)00534-2](https://doi.org/10.1016/s0092-6566(02)00534-2)
- Mendiburo-Seguel, A., Paez, D., & Martinez-Sanchez, F. (2015). Humor styles and personality: A meta-analysis of the relation between humor styles and the Big Five personality traits. *Scandinavian Journal of Psychology, 56*(3), 335–340. <https://doi.org/10.1111/sjop.12209>
- Nunnally, J. C. (1967). *Psychometric theory*. McGraw-Hill.
- Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2017). Ha ha? Assessing individual differences in humor production ability. *Psychology of Aesthetics, Creativity, and the Arts, 11*(2), 231–241. <https://doi.org/10.1037/aca0000086>
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>
- Plesken, C. Y., Franken, F. R., Ster, C., Schmid, R. R., Wolfmayr, C., Mayer, A.-M., ... Tran, U. S. (2020). Humor styles and personality: A systematic review and meta-analysis on the relations between humor styles and the Big Five personality traits. *Personality and Individual Differences, 154*. <https://doi.org/10.1016/j.paid.2019.109676>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Risch, A. K., Strohmayr, C., & Stangier, U. (2005). Fragebogen zum Psychologischen Wohlbefinden [Questionnaire on psychological well-being]. In F. S. University (Ed.), *Jena, Germany*.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. <http://www.jstatsoft.org/v48/i02/>
- Ruch, W., & Heintz, S. (2016). The German version of the Humor Styles Questionnaire: Psychometric properties and overlap with other styles of humor. *Europe's Journal of Psychology, 12*(3), 434–455. <https://doi.org/10.5964/ejop.v12i3.1116>
- Ruch, W., & Heintz, S. (2017). Experimentally manipulating items informs on the (limited) construct and criterion validity of the Humor Styles Questionnaire. *Frontiers in Psychology, 8*, 616. <https://doi.org/10.3389/fpsyg.2017.00616>
- Ryff, C. D., & Keyes, C. L. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719–727. <https://doi.org/10.1037//0022-3514.69.4.719>
- Scheel, T., Gerdenitsch, C., & Korunka, C. (2016). Humor at work: Validation of the short work-related Humor Styles Questionnaire (swHSQ). *Humor, 29*(3). <https://doi.org/10.1515/humor-2015-0118>
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*(6), 1063–1078. <https://doi.org/10.1037//0022-3514.67.6.1063>
- Schermer, J. A., Rogoza, R., Kwiatkowska, M. M., Kowalski, C. M., Aquino, S., Ardi, R., ... Krammer, G. (2019). Humor styles across 28 countries. *Current Psychology, 18*. <https://doi.org/10.1007/s12144-019-00552-y>
- Schneider, M., Voracek, M., & Tran, U. S. (2018). "A joke a day keeps the doctor away?" Meta-analytical evidence of differential associations of habitual humor styles with mental health. *Scandinavian Journal of Psychology, 59*(3), 289–300. <https://doi.org/10.1111/sjop.12432>
- Schönbrodt, F. D., & Gerstenberg, F. X. R. (2012). An IRT analysis of motive questionnaires: The Unified Motive Scales. *Journal of Research in Personality, 46*(6), 725–742. <https://doi.org/10.1016/j.jrp.2012.08.010>
- Schupp, J., & Gerlitz, J.-Y. (2008). Big Five Inventory-SOEP (BFI-S). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. <https://doi.org/10.6102/zis54>
- Silvia, P. J., & Rodriguez, R. M. (2020). Time to renovate the Humor Styles Questionnaire? An item response theory analysis of the HSQ. *Behavioral Sciences, 10* (11), 173. <https://doi.org/10.3390/bs10110173>
- Sirigatti, S., Penzo, I., Giannetti, E., & Stefanile, C. (2014). The Humor Styles Questionnaire in Italy: Psychometric properties and relationships with psychological well-being. *Europe's Journal of Psychology, 10*(3), 429–450. <https://doi.org/10.5964/ejop.v10i3.682>
- Sulejmanov, F., & Renner, K.-H. (2017). Development and initial validation of the Single-Item Wittiness Scale: A preliminary study. *Israeli Journal for Humor Research, 1*, 54–66. <http://hdl.handle.net/20.500.12188/964>
- Torres-Marín, J., Navarro-Carrillo, G., & Carretero-Dios, H. (2018). Is the use of humor associated with anger management? The assessment of individual differences in humor styles in Spain. *Personality and Individual Differences, 120*, 193–201. <https://doi.org/10.1016/j.paid.2017.08.040>
- von Collani, G., & Werner, R. (2005). Self-related and motivational constructs as determinants of aggression. *Personality and Individual Differences, 38*(7), 1631–1643. <https://doi.org/10.1016/j.paid.2004.09.027>
- Watson, D., & Clark, L. A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. <https://doi.org/10.1037//0022-3514.54.6.1063>
- Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology, 61* (4), 590–597. <https://doi.org/10.1037/0022-3514.61.4.590>

DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

ub | universitäts
bibliothek

Dieser Text wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

DOI: 10.1016/j.actpsy.2024.104141

URN: urn:nbn:de:hbz:465-20250107-141240-5



Dieses Werk kann unter einer Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 Lizenz (CC BY-NC-ND 4.0) genutzt werden.