2022-07

Development and Validation of the Short Version of the Sense of Humor Scale (SHS-S): Paving the Way for Assessment of Humor Skills Training

Sonja Heintz[12], Willibald Ruch[1], Chloe Lau[3], Donald H. Saklofske[3], Paul McGhee[4]

[1]Department of Psychology, University of Zurich, Switzerland

[2]School of Psychology, University of Plymouth, United Kingdom

[3]Psychology Department, University of Western Ontario, Canada

[4]The Laughter Remedy, Wilmington, DE, USA

Correspondence concerning this article should be addressed to: Dr. Sonja Heintz, School of Psychology, Faculty of Health, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, United Kingdom. E-mail: sonja.heintz@plymouth.ac.uk

Acknowledgements

We thank Olenka Dworakowski for her assistance with the data analyses.

This manuscript was published as:

Heintz, S., Ruch, W., Lau, C., Saklofske, D. H., & McGhee, P. E. (2022). Development and validation of the short version of the Sense of Humor Scale (SHS-S): Paving the way for assessment of humor skills training. *European Journal of Psychological Assessment, 38*(4), 320–331. https://doi.org/10.1027/1015-5759/a000670

Abstract

Humor training has become increasingly popular to enhance the 'sense of humor' and well-being and to decrease depressive symptoms. Despite the wide applications of these training programs, the assessment of training efficacy has attracted less attention. The *Sense of Humor Scale* (SHS; McGhee, 1996, 1999) recently was expanded to a long version (SHS-L) to enhance its internal consistency (Ruch & Heintz, 2018). At the same time, there is also the need for a brief version of this scale. The purpose of the present study is to develop a short version (SHS-S) in both German- and English-speaking countries, ~~to~~ test its psychometric properties (internal consistency, factorial, construct, and criterion validity), and assess measurement invariance across gender and the two languages. Using three samples (Sample 1: 570 English-speakers, Sample 2: 353 German-speakers, Sample 3: 94 other-reports), the 29-item SHS-S was developed and yielded promising internal consistency and validity scores for the six humor skill factor~~s~~ of enjoyment of humor, laughter, verbal humor, finding humor in everyday life, laughing at yourself, and humor under stress. Overall, the SHS-S is an internally consistent, valid, and economic tool for future research and group-based applications, while the SHS-L seems especially useful in individual applications.

Keywords: Sense of humor; Validation; Humor skills; Measurement invariance; Humor training

The relevance and relationship of humor to both psychological and physical health and well-being has been extensively described in the research literature (for overviews, see Ruch & Hofmann, 2017; Ruch & McGhee, 2014; Schneider, Voracek, & Tran, 2018). Over the past several decades, increasing attention has been directed at humor-based training and interventions in clinical settings (e.g., Hirsch et al., 2010), and the workplace (e.g., McGhee, 2010; Rocke, 2015) as well as in the positive psychology literature (e.g., Wellenzohn et al., 2016). The most widely used and supported training is McGhee's Seven Humor Habits Program, initially published in 1996, and revised and updated in 1999 and 2010. This training program aims to foster the basis of the sense of humor (playfulness and positive mood) as well as development of six humor skills (enjoyment of humor, laughter, verbal humor, finding humor in everyday life, laughing at yourself, and humor under stress). The *Sense of Humor Scale* (SHS) was developed to assess the training outcomes. After an initial appraisal in a German and American sample (Ruch & Carrell, 1998), this scale underwent a revision (McGhee, 1999, 2010) and later an extension (Ruch & Heintz, 2018). The aims of the present study are to develop and test the psychometric properties of the long (SHS-L) and short form (SHS-S) of the SHS.

The Six Humor Skills

Sense of humor can be seen as either a narrow and positive trait or as a multidimensional construct comprising all humor-related individual differences (Ruch & Heintz, 2019). The six humor skills fall in the first category, as these attributes are expected to relate to positive outcomes and can be combined to yield a total sense of humor score (McGhee, 1996, 1999, 2010). Further studies showed however that separating the six humor skills could provide additional insights beyond the total score. For example, Ruch and Heintz

(2018) found that each of the six humor skills showed unique relationships to the temperamental basis of the sense of humor and specific humor factors, while all skills were positively related to life satisfaction. Additionally, the individual six humor skills explained more variance in the humor-outcomes and life satisfaction than the total scale.

It is not surprising that these skills relate to different aspects and styles of the sense of humor. Although all of them are positively connoted, they capture different humor domains (see Table 1). Three general humor domains that are distinguished in humor research are comprehension (i.e., how humor is understood and interpreted), appreciation (i.e., how funny, boring, or offensive a humorous stimuli is evaluated), and production (i.e., how much and which type of humor is created or reproduced). Furthermore, we have listed similar humor-related constructs that are typically investigated in psychological humor research to outline the conceptual nomological network of the six humor skills.

The first skill, enjoyment of humor, reflects a passive appreciation of humor. It comprises seeking out humorous stimuli (such as cartoons or funny videos) and is meant to provide an easy start into humor training. The next skill, laughter, refers to 'hearty belly laughter'. Although laughter is not specific to humor, this skill also emphasizes the attitude underlying laughter: letting go of constraints, feeling lightness, and enjoying the moment. The third skill, verbal humor, broadly focuses on humor production. It ranges from the reproduction of 'canned' humor (such as learning jokes by heart and practicing them) to creating spontaneous word plays and being witty on the spot. The next skill, finding humor in everyday life, mostly covers humor detection and appreciation, in contrast to enjoyment of humor. It requires finding amusing sides in situations and aspects that are not inherently funny and thus requires reframing one's perception from a serious to a playful interpretation. The fifth skill, laughing at yourself, combines humor detection, appreciation, and production related to oneself. It incorporates detecting aspects of oneself that can be construed as funny,

being amused by these aspects, and finally making fun of these aspects. The final skill, humor under stress, refers to detecting humor in stressful situations and conditions, reframing them in a funny way, and laughing and joking about them. The humor skills are ordered from least to more difficult to facilitate during the training process. As can been seen in Table 1, the more difficult skills (4–6) indeed tap into more humor domains than the simpler ones (1–3).

The Need for a Revised Scale

Separating the skills and relating them to other humor-related traits (as compiled in Table 1) can provide insights into the similarities and differences among these constructs and thus help to prevent jingle-jangle fallacies (Block, 2000). For instance, does finding humor in everyday life correspond to benevolent humor (Ruch & Heintz, 2016)? Or does humor under stress correspond to coping humor (Martin & Lefcourt, 1983)? Reliable and valid scales allow for integrating the six humor skills into the broader literature on humor traits and support the work for developing comprehensive models of individual differences in humor (see also Ruch & Heintz, 2019).

The main utility of the SHS is in applied research and applications that employ the Seven Humor Habits Training programs developed by (McGhee, 2010). This program is comprised of the following seven habits, which are trained in an 8-week program: (1) Characterizing one's sense of humor, (2) Becoming less serious and more playful, (3) Working on one's belly laugh, (4) Improving one's ability to tell jokes and create spontaneous humor, (5) Finding humor in daily life, (6) Learning to laugh at oneself, (7) Using these humor habits to cope with stress. The six humor skills assessed in the SHS are associated with habits 1 and 3-7 of the program (habit 2 is assessed with two additional scales of mood and playfulness). Most research only investigated, and supported the notion, that the total SHS score can be increased by the Seven Humor Habits Training (for an overview, see Ruch & McGhee, 2014). As each of the six skills is trained in the program, it would be relevant to

delineate, their differential response to training. Having a short, reliable, and valid measurement instrument for the six humor skills would allow for further research on the effectiveness of training and for evaluations of humor trainings in a variety of applied settings. and answers to several basic and applied questions: Which humor skills are fostered in the humor training? Do pre-existing trait levels of the humor skills moderate the effectiveness of the humor training? Do post-training increases in certain skills mediate the positive outcomes of the humor training? Does training the individual habits increase the levels of this specific skill? Are there interactive effects among the humor skills in influencing training success?

There was support for a one-factor structure (Ruch & Carrell 1998; Wrench & McCroskey, 2001) of the original (McGhee, 1996) and revised SHS (McGhee, ~~1996,~~ 1999). As psychometric investigations yielded insufficient internal consistency for some facets (laughter and enjoyment of humor subscales), it was recommended to use the total score only (Proyer et al., 2010; Ruch & Carrell, 1998; Wrench & McCroskey 2001). To our knowledge, construct validity has only been tested thus far for the laughing at yourself subscale, which supported its convergent validity with peer-reports and observed behavior (Beermann & Ruch, 2011; Hofmann, 2018). A recent extension constructed a parallel form of the SHS, doubling the number of items for each skill from four to eight (Ruch & Heintz, 2018). Combining the SHS and its parallel version yielded satisfactory internal consistencies for all six humor skill subscales (Cronbach's alpha .76 to .96). Furthermore, CFA supported acceptable fits of different models that distinguish the six humor skills (six-factor, hierarchical and bifactor models), while a one-factor model was not supported. They also excluded one item (Item 19 from the enjoyment of humor subscale) as it showed higher correlations with the other five subscales than with the skill it was assigned to.

The Present Study

First, we report the internal consistency and validity of the English and German long version of the SHS (SHS-L, with 48 items in total) based on observed scale scores. The second goal is to select the optimal performing items from the (SHS-L) to create a short version, the SHS-S. Third, we test the internal consistency and validity of the six SHS-S scale scores; specifically, we investigate factorial validity (via exploratory structural equation modeling and confirmatory factor analyses), criterion validity with life satisfaction and construct validity (convergent and discriminant validity) with other-reports using congeneric factor models. Life satisfaction was chosen as a criterion as the humor training not only aims at fostering the sense of humor, but also wellbeing more generally (McGhee, 2010), and life satisfaction is a standard outcome against which the SHS has been tested (e.g. Ruch & Heintz, 2018; Ruch, Hofmann, Rusch, & Stolz, 2018). Finally, measurement invariance of the SHS-S scales is tested to determine the extent to which the scales allow for comparisons across gender and language (English- and German-speaking). This procedure should result in an economic, reliable, and valid assessment instrument of the six humor skills that can be employed in both German-speaking and English-speaking countries–the SHS-S. The present study thus examines the following research questions:

Research Question 1: Can the internal consistency and validity of the six SHS-L sum scores be supported?

Research Question 2: What is the optimal combination of items from the SHS-L to allow an economic, internally consistent, and factor-valid assessment of the six humor skills factors in German- and English-speaking countries (SHS-S)?

Research Question 3: Can the internal consistency and validity of the six SHS-S factor scores be supported?

Research Question 4: Is the SHS-S factor model measurement invariant across gender and language (English vs. German)?

Materials and Methods

Participants

Sample 1. Sample 1 was comprised of English-speakers attending a Canadian university who volunteered to complete self-reports of the SHS-L and the SWLS. Of the 615 students who started the questionnaires, 3 were excluded because they did not complete all SHS-L items, 40 were excluded because they failed to pass an attention check item ("Please choose Neutral"), and 3 participants were excluded because they always chose the same response option in the SHS and the parallel version. This left a total of 570 valid responders with a mean age of 18.38 years old (*SD* = 1.78, range 17–38), and who were mostly female (72%, 27% male, 1% other/prefer not to say), and Canadian (81%, 19% with other nationality).

Sample 2. Sample 2 consisted of participants from German-speaking Switzerland and Germany who completed self-reports of the SHS-L and the SWLS. Of the 381 participants who started the questionnaires, 15 were excluded because they did not complete all SHS-L items, 2 were excluded because they indicated an age of less than 18 years, and 11 participants were excluded because they responded too quickly (i.e., answered >20 items/min). This left a total of 353 valid responders in Sample 2. They were on average 25.64 years old (*SD* = 10.65, range 18–92), most of them were female (78%, 22% male), and Swiss (69%; 24% German, 7% with other nationality).

Sample 3. Sample 3 consisted of participants from German-speaking Switzerland and Germany who completed other-reports of the SHS and the parallel version. A total of 170 participants rated the participants in Sample 2 on the same items (rephrased in third-person). Specifically, participants were asked to add the first name (or nickname) of the person to be rated as well as their gender, which was then automatically added and adapted in the items.

For example, the item “I often laugh till I get tears in the eyes” would have read “X often laughs till he/she gets tears in the eyes”. If more than one person rated a target (range 1-3 person per target), these ratings were averaged for each target person to yield other-reports for a total of 94 participants from Sample 2. The raters knew the target person on average for 12.44 years (*SD* = 8.81, range 1-40 years) and indicated they knew them very well (*M* = 6.22, *SD* = 0.95, range 3-7 on a scale from [1] no knowledge to [7] excellent knowledge).

Sample 1 was recruited via the university undergraduate psychology research sample pool, in which participants would receive credit towards their psychology course upon signing up for the study. For Sample 2, selection criteria for participants were an age of 18 years and above and a reasonable command of German. Sample 2 was recruited via university mailing lists, websites, and bulletins, and hence many participants were psychology students of [Author’s University (removed for review)]. Sample 3 was recruited by the participants from Sample 2, who were asked to forward an anonymous link to the other-report survey to two persons known to them. The other-reports were then collected anonymously and the raters were ensured that the target person would not be informed about their ratings. Participants in Samples 2 and 3 could receive a general feedback on the study, collect course credit (for psychology students), and they could participate in a voucher draw.

Sample 1 was employed to develop the SHS-S from the SHS-L. Samples 1 and 2 were employed to test the internal consistency of the SHS-L and the SHS-S scores and the factorial validity and measurement invariance of the SHS-S scores. Samples 2 and 3 were contributed to test the convergent (self-other agreement) and discriminant validity of the SHS-S with other-reports as an independent method from self-reports (see Paunonen & O'Neill, 2010).

Instruments

Sense of Humor Scale (SHS; McGhee 1999). The SHS comprises 24 items to assess the six humor skills with four items each: Enjoyment of humor (sample item “It is important

for me to have a lot of humor in my life"), laughter (sample item "I have a good belly laugh many times each day"), verbal humor (sample item "I often tell jokes"), finding humor in everyday life (sample item "I often share with others the funny incidents I observe, or that happen to me"), laughing at oneself (sample item "I have no trouble poking fun at my physical imperfections"), and humor under stress (sample item "My sense of humor rarely abandons me under stress"). Each item is responded to on a seven-point Likert format ranging from (1) strongly disagree to (7) strongly agree. The items are listed in Supplementary Tables 1 and 2 (https://osf.io/jpcqy/?view_only=df84e3f9402e46789d9a8839f3214ba8).

Parallel form of the Sense of Humor Scale (Ruch & Heintz, 2018). The parallel version of the SHS increased each subscale to eight items, allowing for a more reliable assessment of the six humor skills. Convergent validity with the original SHS (McGhee, 1999) and internal consistency, both of the parallel subscales and the combined scales, was supported. The same Likert response format was used as with the SHS. The SHS and the parallel version together form the 48-item SHS-L. The items are listed in Supplementary Tables 1 and 2 (https://osf.io/jpcqy/?view_only=df84e3f9402e46789d9a8839f3214ba8).

Satisfaction with Life Scale (SWLS; Diener et al., 1985). The SWLS assesses general satisfaction with one's life with five items rated on a seven-point Likert format ranging from (1) strongly disagree to (7) strongly agree.

Procedure

The English version of the parallel version of the SHS was created using a standardized translation-back-translation procedure (see Van de Vijver & Hambleton, 1996). Data collection was conducted in line with the local ethical guidelines (for Sample 1: Non-Medical Research Ethics Board of [Author's University (removed for review)]; for Sample 2 and 3: Ethics Committee of the Faculty of Arts and Social Sciences of [Author's University (removed for review)] and received ethical approval where needed. All samples

were collected online. Other variables were assessed (comic styles, cheerfulness, humor factors) that are not relevant for the present study. Samples 1 and 2 partially overlap with a previously published study (AUTHORS, 2019); however, none of the present findings have been reported previously.

Data analysis

All analyses were conducted in *R* (R Core Team 2020) using the packages *lavaan* (Rosseel 2012), *psych* (Revelle, 2019), *ccpsyc* (Karl, 2020), *semTools* (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2020), *lsr* (Navarro, 2015), and *semPlot* (Epskamp, 2019). 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 (analyses and models of previous versions of this manuscript included one-factor and bifactor CFA models, CFA models for unidimensionality, and analyses of item bias). If we use inferential tests, we report exact p values, effect sizes, and 95% confidence or credible intervals.

Evaluation of the SHS-L: Observed mean scores for the six scales were computed. Cronbach's alpha served to test the internal consistency of the SHS-L scale scores. Observed correlations between self-reports (Sample 2) and corresponding other-reports (Sample 3) were computed. The resulting multitrait-multimethod matrix (Campbell, & Fiske, 1959) supports convergent validity if the self-other convergence is large (i.e., $r \geq .30$; Gignac & Szodorai, 2016). Discriminant validity is supported if self-other convergence is numerically larger than the correlations with the other scales in either self- or other-reports.

Construction of the SHS-S. Item selection was based on three steps: (1) Content-based item selection, (2) exploratory structural equation modeling (ESEM; Sample 1), and (3) confirmatory factor analyses (CFA; Sample 2). In Step 1, we deleted items that could not be

generalized or that were not applicable to all participants, that were very similar (only one of the items was retained in this case), and deleted Item 19, which was found to perform poorly in the previous SHS-L study (Ruch & Heintz, 2018). In Step 2, we conducted ESEM for each pairwise combination of factors using the *psych esem* function, an extension of the exploratory factor analysis function *fa*. Items were excluded if (a) their standardized loading was < .40 on the latent factor (see Stevens 2012) or (b) their second loading was higher than their main loading. Overall, this procedure should yield factor-valid, yet shorter scales. In Step 3, we conducted a CFA with six correlated latent factors on the chosen items to cross-validate and confirm the suitability of the item selection for the SHS-S from Step 2. It should be noted that items were not chosen to maximize internal consistency (as the SHS-L would be most suitable for this purpose), but to provide the best selection of items that represents the six humor skills validly and economically.

Internal consistency and validity of the SHS-S. McDonald's omega total served as an indicator of the internal consistency of the scales. Factorial validity was tested in a CFA using a six-factor model, with six correlated humor skill factors). Construct validity (convergent and discriminant validity) was tested by correlating the latent SHS-S scale scores with the latent other-report scores using SEM. This was done separately for each other-reported scale to not make the model too complex, given the small sample size of Sample 3. Criterion validity was tested by correlating the latent SHS-S scale scores with the latent SWLS score using SEM.

Measurement invariance. Measurement invariance was tested for the SHS-S in multi-group CFAs for gender (males and females) and sample/language (German and English). Metric invariance was tested by forcing all item loadings to be equal across groups. This model was then compared with the baseline model that allows a free estimation of the item loadings (configural model), comparing the difference in the CFI and the RMSEA.

Changes of ≤ |.01| and ≤ |.015| in CFI and RMSEA, respectively, were used as cut-offs to indicate measurement invariance (Chen, 2007; Cheung & Rensvold, 1999). Scalar invariance was tested by forcing both the intercepts and the loadings to be equal across groups. Next, residual invariance was tested by forcing the intercepts, loadings, and residual variances to be equal across groups. Finally, strict measurement invariance was tested by also forcing the latent means to be equal across groups.

Results

Internal consistency and validity of the SHS-L

The descriptives of all items in the three samples as well as separate for gender are listed in Supplementary Table S4 (https://osf.io/jpcqy/?view_only=df84e3f9402e46789d9a8839f3214ba8). Table 2 shows the internal consistencies (Cronbach's alpha) of the six SHS-L scores, separate for Samples 1 and 2 and for men and women. Enjoyment of humor had the lowest values in all samples (range .65-.71), while humor under stress had the highest values (range .87-.92). Laughter had higher scores in sample 2 (.81) than in Sample 1 (.69), pointing to a potential influence of language on this factor. All scales except for enjoyment of humor and laughter had Cronbach's alpha values of at least .80.

Table 3 shows the convergent and discriminant validity of the six SHS-L scale scores in a multitrait-multimethod correlation matrix. Self-other convergence exceeded .30 (range .33-.53), indicating a large convergent validity. Furthermore, these convergent correlations were always higher than the discriminant correlations (i.e., correlations with the other scales). Table 3 also shows the correlations of the six SHS-L scales with the SWLS. All correlations, except for enjoyment of humor, were significant. Effects were small for verbal humor and humor under stress, medium-sized for finding humor in everyday life and laughing at yourself, and large for laughter (based on the criteria by Gignac & Szodorai, 2016). Overall,

construct validity was supported for all scales and criterion validity for all scales except for enjoyment of humor.

Item selection for the SHS-S

The 48 items of the original and parallel version of the SHS served as a starting point for the item selection. The descriptives of all items in the three samples as well as separate for gender are listed in Supplementary Table S4 (https://osf.io/jpcqy/?view_only=df84e3f9402e46789d9a8839f3214ba8). The first step in the construction of the SHS-S was content-based item selection. Item 19 (from enjoyment of humor) was excluded as it was found to have insufficient psychometric properties in a previous study (Ruch & Heintz, 2018). Item 7 (enjoyment), 20 (laughter), 4 and 16 (finding humor in everyday life), and 18 (humor under stress) were excluded because they were deemed too specific, which could make it difficult to answer the item;~~.~~ they referred to specific contexts (work, settings outside of work and family), specific activities (looking for cartoons in magazines), or past events (parents' laughter during childhood). Furthermore, four items were excluded due to their similarity with other items: Items 37 (similar with Item 1), 24 (similar to Item 6), 17 (similar to Item 11), and 21 (similar to Item 27). While this might reduce internal consistency, the item contents were sufficiently redundant to remove the items without changing the meaning of the scale score (content validity). This first step le~~a~~d to the exclusion of 10 of the 48 SHS-L items.

In the second step, we conducted ESEM in Sample 1 to for each pairwise combination of factors. Items were excluded if (a) their standardized loading was < .40 on the latent factor or (b) their second loading was higher than their main loading. This resulted in the deletion of items 9 items: Items 25 and 43 (enjoyment), 14, 26, 32, and 44 (laughter), 45 (verbal humor) 41 (laughing at yourself), and 42 (humor under stress). This step resulting in a total of 29

items for the SHS-S, 4 for enjoyment of humor, 3 for laughter, 6 each for verbal humor and finding humor in everyday life, and 5 each for laughing at yourself and humor under stress.

To cross-validate and confirm these findings in an independent sample, we conducted a CFA with six correlated latent factors on the 29 items in a third step. The resulting model is shown in Figure 1. All loadings were > .40 and significant at $p < .001$. The model fit was $\chi^2 = 996$, $df = 362$, $\chi^2/df = 2.75$ (acceptable), comparative fit index (*CFI*) = .873 (insufficient), root mean square error of approximation (*RMSEA*) = .070 (90% CI [.065, .076]; acceptable), and standardized root mean square residual (*SRMR*) = .061 (acceptable) based on Schermelleh-Engel, Moosbrugger, and Müller's (2003) guidelines. Thus, except for the CFI, all fit indices indicate an acceptable fit. Together with the large and significant loadings, the model can thus be considered replicated in Sample 2. As expected, the six humor skills showed large latent correlations, ranging from .36 (enjoyment of humor and laughing at yourself) to .85 (verbal humor and finding humor in everyday life), with a median of .51. Thus, the six scores shared between 13% and 72% of their true-score variance, with a median overlap of 26%.

Internal consistency and validity of the SHS-S

Table 2 shows the internal consistency (McDonald's omega total) of the six SHS-S scales. Similar to the SHS-L, the internal consistencies of enjoyment of humor and laughter were below .70 in most samples, while all other scales had values > .80. Enjoyment of humor had again the lowest and humor under stress the highest scores. Latent self-other correlations of the SHS-S scales with the other-reports are shown in Table 3. Self-other convergence always exceeded .30 (range .35-.74), supporting the convergent validity of the six SHS-S scores. Also, these latent convergent correlations were higher than the latent discriminant correlations with the other scale scores. Criterion validity of the scale scores was also supported by significant and small to medium-sized positive correlations with the SWLS, except for

enjoyment of humor. Overall, these findings support the construct validity of all SHS-S scores and their criterion validity, with the exception of enjoyment of humor.

Measurement invariance

Measurement invariance of the SHS-S across samples (English- and German-speaking) and gender (men and women) was tested using multi-group CFA models (visualized in Figure 1). The fit indices of the multi-group different models and the change between each subsequent restriction (loadings, intercepts, residuals, and means) are shown in Table 4. The χ^2 change were always significant, as can be expected in the large samples. The CFI change was > |.01| for the scalar invariance models for both models, indicating that the intercepts differ across samples and gender. The RMSEA change, however, supported strict measurement invariance, including intercepts, residuals and means, as all changes were ≤ |.015|. Thus, the SHS-S scores should only be compared with caution across the two languages and gender due to the inconsistency in these findings.

Discussion

The objectives of the present study included (1) testing the internal consistency and validity of the SHS-L sum scores, (2) developing a short version of the SHS-L (the SHS-S), (3) testing the internal consistency and validity of the SHS-S factor scores, and (4) testing the measurement invariance of the SHS-S scales across gender and language (English vs. German). Regarding the first aim, the observed SHS-L sum scores were found to be internally consistent and valid, with the exception of the enjoyment of humor scale. Regarding the second aim, we developed short version with 29 of the original 48 items. Regarding the third aim, The latent SHS-S factors were found to be internally consistent and valid, again with the exception of the enjoyment of humor scale. Regarding the fourth aim, measurement invariance of the SHS-S scales was not fully across gender and languages.

Focusing on selecting items that enhance brevity while supporting both reliability and validity, 29 items were selected from the original 48 items that allow the best combination to assess the six humor skills in both German and English. The SHS-S scales had three to six items for each humor skill, allowing a short assessment of the six humor scales. Internal consistency of the six scale scores (Cronbach's alpha and, McDonald's omega total) was mostly supported for both the SHS-L and the SHS-S, with enjoyment of humor and laughter yielding the lowest reliabilities. These scores were especially low in Sample 1, pointing to a greater heterogeneity of the interpretations of these items in English than in German. Still, all scores are sufficient for group-based analyses. Four of the scale scores (verbal humor, finding humor in everyday life, laughing at yourself, and humor under stress) also yielded sufficient internal consistencies to allow comparisons at the person-level (e.g., assessing one person before and after a humor training). The SHS-L can be recommended for individual purposes, as its internal consistency and validity was supported in observed sum scores. The SHS-S is most suitable for group-based applications and research purposes with sufficient sample sizes to compute the six-factor model.

The present study is the first to investigate the construct validity of the SHS. Comparing self- and other- reports of the SHS-L and SHS-S scales, both convergent and discriminant validity were supported. Hence, others' perceptions of a person's humor skills converged with the self-description of this person, supporting the "social reality" of the SHS scale scores. The range of self-other convergence was similar to other personality traits (~.30-.6070), and higher in humor skills that can be easily observed by others (e.g., verbal humor, laughter) than in those that are more introspective (finding humor in everyday life, laughing at yourself, humor under stress). In addition to supporting the SHS-L and SHS-S scores' construct validity, this also suggests that observer ratings can offer a valid independent method of assessing an individual's training success (as implemented by Ruch et al., 2018).

Finally, correlating the SHS scales with life satisfaction showed the expected pattern of positive correlations, both for the SHS-L (replicating Ruch & Heintz, 2018) and for the SHS-S. Thus, the SHS-S retained the positive aspects of the sense of humor and the six humor skills. Only enjoyment of humor yielded non-significant and negligible correlations with life satisfaction, which is however similar to previous findings (Ruch & Heintz, 2018).

Another novel contribution of the present study is testing the measurement invariance of the SHS-S across two languages and gender. The support for measurement invariance was inconsistent, with the CFI change indicating measurement variance and the RMSEA change indicating strong measurement invariance. The present data cannot reveal why measurement invariance was not supported, and further research is needed to better understand the generalizability of the factorial model of the SHS-S. In addition to the language, which might have made some items "easier" or more desirable in one language than another, the samples also differed in other characteristics. For example, the English-speaking sample was comprised of young college students, while the German-speaking sample was drawn both from both university students and the general population. Additionally, participants in both samples had different cultural backgrounds, which might have influenced the results (e.g., Heintz et al., 2018; Schermer & Kfrerer, 2020). Thus, future studies should disentangle these potentially different sources of variance by recruiting representative or comparable samples from different languages. This would allow for a better understanding of how the sense of humor is shaped by environmental factors.

Practical implications of the inconsistent support for the SHS-S's measurement invariance means that studies that include both language versions of the instrument might find differences that could be due to methodological (e.g., translation, item interpretation, differentiation between the six scales), rather than trait related effects (i.e., genuine differences in the sense of humor scores). The same holds for gender differences in the six

humor skills. While gender differences are often investigated in humor (Hofmann, Platt, Lau, & Torres-Marín, 2020), very few studies also tested whether the scores in question are measurement invariant and whether they can be meaningfully interpreted. Conceptually, however, we would not expect large gender differences in any of the SHS-S scales, as men were found to score higher than women in aggressive styles of humor, while the more benevolent forms assessed by the SHS-S are usually similar across gender (Hofmann et al., 2020). Nonetheless, the nature of measurement variance and invariance should be investigated further, both for improving humor assessment and applications.

Overall, the SHS-S and SHS-L can be recommended as internally consistent, construct-valid and criterion-valid latent scales for assessing the six humor skills defined as enjoyment of humor, laughter, verbal humor, finding humor in everyday life, laughing at yourself, and humor under stress. Due to its brevity, the SHS-S is particularly suitable for research in which group-based data are of interest and for applied settings in which participant burden should be reduced (e.g., patients, clinical groups) and in which time constraints apply. The SHS-S items are also generalizable to samples from varied backgrounds, as items referring to specific contexts or activities were excluded. Furthermore, it should be kept in mind that the psychometric properties SHS-S scores only apply to congeneric factor models, and not to observed sum scores (unless constraints are applied; see McNeish & Wolf, 2020). If individuals are to be compared on enjoyment of humor and laughter, the SHS-L is recommended due to its higher internal consistency. Furthermore, if the observed sum scores are of interest, the SHS-L should be used, as the psychometric properties were supported for its sum scores. The present study thus provides a contribution both for research, in which a validated assessment of the six humor skills has been lacking thus far, and for evaluations of humor trainings in a variety of settings.

Limitations and suggestions for future research

Both samples were convenience samples, in which young adults, females and well-educated participants were overrepresented. Testing the psychometric properties of the SHS-L and the SHS-S in more varied samples would be desirable to test the generalizability of the present findings and to determine specific groups for which further adaptations might be warranted. While the self-reports were comprised of large samples, the other-report sample (Sample ~~2~~ 3) was small (94 participants with matched self-reports) and hence self-other agreement might have been underestimated. This also prevented us from using more sophisticated more sophisticated SEM-based analyses (such as the multiple indicator CT-C(M-1) model; Eid, Lischetzke, & Nussbeck, 2006; Geiser, & Simmons, 2021) or Bayesian models (e.g., Helm, Castro-Schilo & Oravecz, 2017). Replicating the construct validity of the SHS-L and SHS-S in larger samples, different languages, and more sophisticated models would be important directions for future research. Furthermore, given the application for humor trainings, studies on the test-retest reliability and change sensitivity would be important next steps to better understand the temporal dynamics underlying the six humor skills. Finally, as the internal consistency and validity of enjoyment of humor was the lowest, it could be advisable to add further items specific to the sample of interest if a more internally consistent is desired; for example, the modalities (such as Internet, books, magazines, TV) could be adapted to make the items relevant for the specific sample.

Conclusions

The short and long version of the Sense of Humor Scale can be recommended as psychometrically sound and economic instruments for humor research studies and to assess training success of McGhee's Seven Humor Habits Program (7HHP; McGhee, 2010). The scales also allow future cross-cultural research to further improve the assessment and training of humor and to widen it to new applied settings. At the same time, these measures contribute

to a cumulative science of the six humor skills, which can be integrated into the larger endeavor of comprehensively mapping the sense of humor (Ruch & Heintz, 2019).

Open Science

Open Data: The information needed to reproduce all of the reported results are 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.

Supplementary materials for this manuscript is available at https://osf.io/jpcqy/?view_only=df84e3f9402e46789d9a8839f3214ba8

References

AUTHORS, 2019

Beermann, U. & Ruch, W. (2011). How virtuous is humor? What we can learn from current instruments. *The Journal of Positive Psychology, 4*, 528–539. https://doi.org/10.1080/17439760903262859

Block, J. (2000). Three tasks for personality psychology. In L.R. Bergman, R.B. Cairns, L.-G. Nilsson, & L. Nystedt (Eds.), *Developmental science and the holistic approach* (pp. 155–164). Mahwah, NJ: Erlbaum.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. https://doi.org/10.1037/h0046016

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G.W., & Rensvold, R.B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management, 25*, 1–27. https://doi.org/10.1177/014920639902500101

Diener, E., Emmons, R.A., Larsen, R.J. & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Eid, M., Lischetzke, T., & Nussbeck, F.W. (2006). Structural Equation Models for Multitrait-Multimethod Data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: American Psychological Association.

Epskamp, S. (2019). *semPlot: Path diagrams and visual analysis of various SEM packages' output.* R package version 1.1.2. https://CRAN.R-project.org/package=semPlot

Geiser, C., & Simmons, T.G. (2021). Do method effects generalize across traits (and what if they don't)? *Journal of Personality.* Advance online publication. https://doi.org/10.1111/jopy.12625

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Helm, J.L., Castro-Schilo, L., & Oravecz, Z. (2017). Bayesian versus maximum likelihood estimation of multitrait–multimethod confirmatory factor models. *Structural Equation Modeling*, *24*, 17-30. https://doi.org/10.1080/10705511.2016.1236261

Hirsch, R.D., Junglas, K., Konradt, B., & Jonitz, M.F. (2010). Humor therapy in the depressed elderly: Results of an empirical study. *Zeitschrift fur Gerontologie und Geriatrie*, *43*, 42-52. https://doi.org/10.1007/s00391-009-0086-9

Hofmann, J., Platt, T., Lau, C., & Torres-Marín, J. (2020). Gender differences in humor-related traits, humor appreciation, production, comprehension,(neural) responses, use, and correlates: A systematic review. *Current Psychology.* Advance online publication https://doi.org/10.1007/s12144-020-00724-1

Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A.M., & Rosseel, Y. (2020). *semTools: Useful tools for structural equation modeling.* R package version 0.5-3. https://CRAN.R-project.org/package=semTools

Martin, R.A., & Lefcourt, H.M. (1983). Sense of humor as a moderator of the relation between stressors and moods. *Journal of Personality and Social Psychology*, *45*, 1313–1324. https://doi.org/10.1037/0022-3514.45.6.1313

McGhee, P.E. (1996). *Health, healing and the amuse system: Humor as survival training.* Dubuque: Kendall/Hunt.

McGhee, P.E. (1999). *Health, healing and the amuse system: Humor as survival training (3rd ed.).* Dubuque: Kendall/Hunt.

McGhee, P.E. (2010). *Humor as survival training for a stressed-out world: The 7 humor habits program.* Bloomington: AuthorHouse.

McNeish, D., & Wolf, M.G. (2020). Thinking twice about sum scores. *Behavior Research Methods, 52,* 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Müller, L. & Ruch, W. (2011). Humor and strengths of character. *The Journal of Positive Psychology, 6.* 368–376. https://doi.org/10.1080/17439760.2011.592508

Navarro, D.J. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners* (Version 0.5). University of Adelaide. Adelaide, Australia.

Paunonen, S.V., & O'Neill, T.A. (2010). Self-reports, peer ratings and construct validity. *European Journal of Personality, 24*(3), 189–206. https://doi.org/10.1002/per.751

Proyer, R.T., Ruch, W. & Müller, L. (2010). Sense of humor among the elderly: Findings with the German version of the SHS. *Zeitschrift für Gerontologie und Geriatrie, 43,* 19–24. https://doi.org/10.1007/s00391-009-0082-0

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Revelle, W. (2019) *psych: Procedures for Personality and Psychological Research.* Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version = 1.9.12

Rocke, C. (2015). The use of humor to help bridge cultural divides: An exploration of a workplace cultural awareness workshop. *Social Work with Groups, 38,* 152–169. https://doi.org/10.1080/01609513.2014.968944

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2). 1–36.

Ruch, W. & Carrell, A. (1998). Trait cheerfulness and the sense of humor. *Personality and Individual Differences, 24*, 551–558.

Ruch, W. & Heintz, S. (2016). The virtue gap in humor: Exploring benevolent and corrective humor. *Translational Issues in Psychological Science, 2*, 35–45.

Ruch, W., & Heintz, S. (2018). Psychometric evaluation of the revised Sense of Humor Scale and the construction of a parallel form. *Humor: International Journal of Humor Research, 31*, 235–257. https://doi.org/10.1515/humor-2016-0085

Ruch, W., & Heintz, S. (2019). On the dimensionality of humorous conduct and associations with humor traits and behaviors. *Humor: International Journal of Humor Research, 32*, 643–666. https://doi.org/10.1515/humor-2018-0119

Ruch, W., & Hofmann, J. (2017). Fostering humour. In C. Proctor (Ed.), *Positive psychology interventions in practice* (pp. 65-80). Cham: Springer.

Ruch, W., Hofmann, J., Rusch, S., & Stolz, H. (2018). Training the sense of humor with the 7 Humor Habits Program and satisfaction with life. *Humor: International Journal of Humor Research, 31*, 287-309. https://doi.org/10.1515/humor-2017-0099

Ruch, W. & McGhee, P.E.. (2014). Humor intervention programs. In A.C. Parks & S.M. Schueller (Eds.), *The Wiley Blackwell Handbook of Positive Psychological Interventions* (pp. 179–193). Chichester: John Wiley & Sons.

Schermer, J.A., & Kfrerer, M.L. (2020). Humor style differences across four English-speaking countries. *Humor: International Journal of Humor Research*. Advance online publication. https://doi.org/10.1515/humor-2019-0027

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.

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,* 289-300. https://doi.org/10.1111/sjop.12432

Stevens, J.P. (2012). *Applied multivariate statistics for the social sciences* (5th Ed.). New York, NY: Routledge.

Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests. *European Psychologist, 1,* 89–99. https://doi.org/10.1027/1016-9040.1.2.89

Wellenzohn, S., Proyer, R.T., & Ruch, W. (2016). Humor-based online positive psychology interventions: A randomized placebo-controlled long-term trial. *The Journal of Positive Psychology, 11*, 584-594. https://doi.org/10.1080/17439760.2015.1137624

Wrench, J.S. & McCroskey, J.C. (2001). A temperamental understanding of humor communication and exhilaratability. *Communication Quarterly 49*, 142–159. https://doi.org/10.1080/01463370109385622

Table 1

Overview of the Humor Domains Tapped by the Six Humor Habits as well as Similar Humor-Related Constructs

Humor habit	Humor domain			Similar constructs
	Comprehension	Appreciation	Production	
1. Enjoyment of humor		X		Humor appreciation
2. Laughter		X		Cheerfulness (CH2, CH4)
3. Verbal humor			X	Affiliative humor style, fun, wit
4. Humor in everyday life	X	X		Benevolent humor
5. Laughing at yourself	X	X	X	Gelotophilia (liking to be laughed at), self-deprecating humor
6. Humor under stress	X	X	X	Coping humor, self-enhancing humor style, cheerfulness (CH3)

Notes. CH2 = low threshold for smiling and laughter, CH3 = composed view of adverse life circumstances, CH4 = broad range of active elicitors of cheerfulness and smiling/laughter.

Table 2

Internal Consistencies of the Long (SHS-L) and Short (SHS-S) Sense of Humor Scales split by Gender and Sample

Scales	Sample 1 (N = 570)	Sample 2 (N = 353)	Men (N = 231)	Women (N = 687)
SHS-L scales (Cronbach's α)				
Enjoyment of humor	.65	.72	.70	.71
Laughter	.69	.81	.73	.74
Verbal humor	.83	.87	.86	.85
Finding humor in everyday life	.85	.88	.88	.86
Laughing at yourself	.87	.88	.87	.88
Humor under stress	.87	.92	.88	.91
SHS-S scales (McDonald's ω_t)				
Enjoyment of humor	.56	.57	.61	.61
Laughter	.63	.73	.65	.67
Verbal humor	.81	.85	.83	.82
Finding humor in everyday life	.80	.85	.85	.83
Laughing at yourself	.85	.85	.86	.85
Humor under stress	.83	.89	.84	.87

Table 3

Validity Correlations of the Long and Short Sense of Humor Scales

Self-report scales	(1 Other-report)			(2 Other-report)			(3 Other-report)			(4 Other-report)		
	r	95% *CI*	*p*	*r*	95% *CI*	*p*	*r*	95% *CI*	*p*	*r*	95% *CI*	*p*
SHS-L (observed)												
(1) Enjoyment of humor	**.47**	[.30,.62]	<.001	.06	[-.14,.26]	.549	.18	[-.02,.37]	.079	.12	[-.08,.32]	.240
(2) Laughter	.16	[-.04,.35]	.115	**.48**	[.31,.62]	<.001	.40	[.21,.56]	<.001	.34	[.15,.51]	<.001
(3) Verbal humor	.14	[-.06,.33]	.178	.02	[-.18,.22]	.842	**.53**	[.37,.66]	<.001	.38	[.19,.54]	<.001
(4) Everyday life	.01	[-.19,.22]	.886	.10	[-.10,.30]	.331	.40	[.21,.56]	<.001	**.40**	[.21,.55]	<.001
(5) Laughing at yourself	.02	[-.18,.22]	.853	.12	[-.08,.32]	.236	.30	[.10,.47]	.004	.29	[.09,.46]	.005
(6) Humor under stress	.05	[-.16,.25]	.662	.10	[-.10,.30]	.336	.25	[.05,.43]	.014	.15	[-.05,.34]	.015
SHS-S (latent)												
(1) Enjoyment of humor	**.69**	[.48,.89]	<.001	.20	[-.09,.50]	.169	.26	[.00,.51]	.048	.19	[-.07,.45]	.152
(2) Laughter	.18	[-.02,.38]	.080	**.74**	[.57,.92]	<.001	.34	[.15,.54]	<.001	.28	[.08,.48]	.005
(3) Verbal humor	.16	[-.07,.38]	.165	.14	[-.12,.39]	.291	**.61**	[.46,.76]	<.001	.42	[.22,.61]	<.001
(4) Everyday life	-.03	[-.26,.20]	.801	.18	[-.07,.44]	.158	.40	[.20,.60]	<.001	**.45**	[.26,.64]	<.001
(5) Laughing at yourself	-.15	[-.37,.08]	.201	.19	[-.06,.44]	.144	.21	[-.01,.43]	.060	.23	[.01,.45]	.039
(6) Humor under stress	.04	[-.18,.26]	.721	.19	[-.05,.44]	.124	.20	[-.01,.42]	.066	.12	[-.11,.34]	.304

(Table 3 is continued)

Table 3 *(continued)*

Self-report scales	(5 Other-report)			(6 Other-report)			Satisfaction with life scale		
	r	95% *CI*	*p*	*r*	95% *CI*	*p*	*r*	95% *CI*	*p*
SHS-L (observed)									
(1) Enjoyment of humor	.18	[.19,.31]	<.001	.17	[-.03,.36]	.096	-.04	[-.10,.03]	.277
(2) Laughter	.17	[-.04,.36]	.112	.25	[.05,.43]	.013	.31	[.25,.37]	<.001
(3) Verbal humor	.07	[-.13,.27]	.476	.30	[.10,.47]	.004	.11	[.05,.18]	<.001
(4) Everyday life	.12	[-.08,.32]	.236	.28	[.09,.46]	.005	.22	[.16,.28]	<.001
(5) Laughing at yourself	**.33**	[.14,.50]	.001	.31	[.11,.48]	.002	.22	[.16,.28]	<.001
(6) Humor under stress	.18	[-.02,.37]	.080	**.34**	[.15,.51]	<.001	.09	[.02,.15]	.009
SHS-S (latent)									
(1) Enjoyment of humor	.20	[-.06,.46]	.125	.08	[-.19,.35]	.565	.03	[-.06,.12]	.496
(2) Laughter	.12	[-.08,.32]	.259	.28	[.08,.48]	.007	.28	[.20,.36]	<.001
(3) Verbal humor	.12	[-.10,.34]	.294	.32	[.11,.53]	.003	.14	[.07,.22]	<.001
(4) Everyday life	.13	[-.10,.35]	.269	.26	[.03,.48]	.024	.27	[.20,.34]	<.001
(5) Laughing at yourself	**.35**	[.15,.55]	<.001	.29	[.08,.51]	.007	.28	[.21,.35]	<.001
(6) Humor under stress	.20	[-.02,.41]	.074	**.37**	[.17,.57]	<.001	.13	[.05,.20]	<.001

Note. *N* = 94 (Samples 2 and 3 for the other-reports), N = 889 for SWLS (Samples 1 and 3). Everyday life = finding humor in everyday life. Self-other convergence (convergent validity) in bold.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Fit Indices of the Measurement Invariance across Samples/Languages of the Short Sense of Humor Scale (SHS-S)

Scales	χ^2	$\Delta\chi^2$	*CFI*	Δ*CFI*	*RMSEA*	Δ*RMSEA*
Sample (*N* = 923)						
M1: Configural	2091	–	.879	–	.064	–
M2: Loadings	2164	74***	.875	.004	.064	.000
M3: Intercepts	2818	654***	.819	.056	.076	.012
M4: Residuals	3008	190***	.805	.014	.077	.001
M5: Means	3268	259***	.782	.022	.081	.004
Gender						
M1: Configural	2196	–	.875	–	.067	–
M2: Loadings	2251	55***	.873	.003	.066	.000
M3: Intercepts	2440	189***	.859	.014	.069	.003
M4: Residuals	2501	62***	.856	.003	.068	.001
M5: Means	2550	49***	.852	.004	.069	.001

Note. *** $p < .001$.

Short Sense of Humor Scale

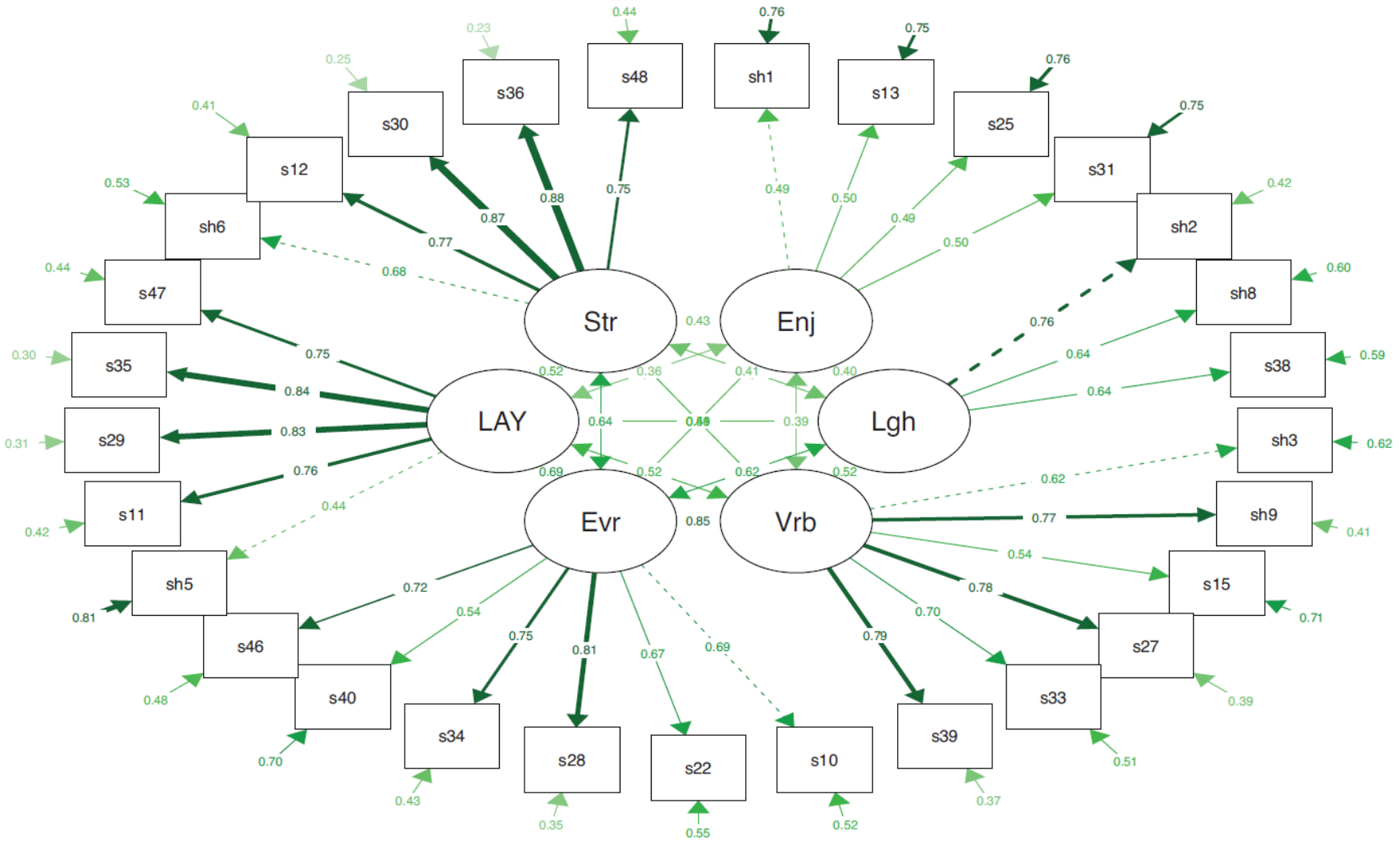

Figure 1. Standardized loadings, residuals and correlations of the Six-Factor Model of the Short Version of the Sense of Humour Scale (Sample 2; N = 353