An examination of the convergence between the conceptualization and the measurement of humor styles: A study of the construct validity of the Humor Styles Questionnaire

Sonja Heintz & Willibald Ruch

University of Zurich, Department of Psychology, Personality and Assessment

Binzmuehlestrasse 14/7, 8050 Zurich, Switzerland

Phone: +41 44 635 7521

E-mail: s.heintz@psychologie.uzh.ch

Abstract

The Humor Styles Questionnaire (HSQ; Martin et al. 2003) was developed using a construct-based scale construction approach to measure four humor styles, namely affiliative, self-enhancing, aggressive, and self-defeating. The present study investigates to what extent the HSQ scales converge with and represent the conceptualizations (i.e., the definitions and construct descriptions) of the four humor styles as outlined by Martin et al. (2003). To this end, 340 participants provided self-reports on the definitions, construct descriptions, and the 32 items of the HSQ. Two multitrait-multimethod analyses yielded a good convergence of the self-defeating humor style, yet for the affiliative, self-enhancing, and aggressive humor styles convergence was lower and they were partly mismatched. The discrimination between the humor styles was mostly supported with the exception of affiliative and self-enhancing. Further, the HSQ scales predicted about two-thirds of the reliable variance in the conceptualizations in multiple regression analyses, so they represented several conceptual elements. Overall, these findings do only lend partial support for the convergence of the HSQ with the original conceptualization of the humor styles. If replicable, this implicates that either the constructs and model of the humor styles need to be adjusted or newly developed, or the HSQ does.

An examination of the convergence between the conceptualization and the measurement of

humor styles: A study of the construct validity of the Humor Styles Questionnaire

1. Introduction

What are the four humor styles sensu Martin and colleagues (2003)? This question has no

simple answer, as there were multiple stages from conceptualizing to measuring the humor styles

from which information can be drawn. Chronologically, these are (1) the literature from which

Martin et al. (2003) derived their 2×2 conceptualization, (2) the definitions of the humor styles

along this conceptualization (pp. 51–52), (3) the construct descriptions of the humor styles (pp. 52–

54), (4) the item pools sampled to measure the humor styles, and (5) the final 32 items measuring

the humor styles (i.e., the Humor Styles Questionnaire / HSQ). This construct-based scale

construction approach (cf. Jackson 1970) should lead to a reliable and construct-valid measurement

of the four hypothesized humor styles constructs. The present study aims at empirically testing the

construct validity of the HSQ by investigating to what extent it converges with two of its

construction stages. Specifically, we test to which degree the HSQ (stage 5 above) converges with

and represents the definitions (stage 2) and construct descriptions (stage 3) of the humor styles.

1.1. *Conceptualizing humor styles*

The HSQ was developed as a self-report measure of everyday functions of humor, especially

those relevant to psychosocial well-being. Martin et al. (2003) derived the four humor styles from

an extensive literature review on the relationship between humor and well-being (stage 1). Their $2 \times$

2 conceptualization of humor functions is supposed to contain "most of the elements discussed in

this literature" (p. 51). One dimension consists of two goals of using humor, to enhance oneself vs.

to enhance relationships with others, which they also referred to as intrapsychic vs. interpersonal

functions of humor. The other dimension consists of two ways of using humor to achieve these

goals, in a benign vs. in a detrimental way. This classification allowed to define four humor styles

(stage 2), i.e., affiliative (enhancing relationships/benign for self), self-enhancing (enhancing

self/benign for others), aggressive (enhancing self/detrimental for others), and self-defeating

(enhancing relationships/detrimental for self). Interestingly, the first two humor styles were named according to their goals, while the latter two were named according to how humor is used to achieve these goals.

The third stage (outlined on pp. 52–54) elaborated on the four humor styles in more detail: "[…] based on this 2×2 model of humor functions, we posit four dimensions relating to individual differences in humor use." (Martin et al. 2003, p. 53). According to Martin et al. (2003), the affiliative humor style entails amusing others and telling jokes. It also involves self-deprecating humor (while still being self-accepting) and gently teasing others (in one's own group). The self-enhancing humor style involves perspective-taking humor and a predisposition to being amused by the incongruities of life. This concept is hypothetically similar to Freud's (1928) theory of humor as a healthy defense mechanism, and it should help to regulate negative affect and to cope. The aggressive humor style entails sarcasm, disparagement humor, teasing, and manipulating others. This humor styles also entails compulsive and impulsive elements. Its theoretical foundation is (among others) Zillmann's (1983) disparagement humor theory, which describes circumstances in which we are amused by humor that ridicules others. The self-defeating humor style mainly refers to disparaging oneself excessively. This humor style should help to become accepted by others (in accord with its interpersonal function), which can also include hiding one's negative affect. It should be noted that this humor style seems not to have been derived from a humor theory or model, but from experience reports.

The process of operationalization is crucial for ensuring that the theoretically derived humor styles are also incorporated into the final questionnaire, the HSQ. Accordingly, Martin et al. (2003) stated "We began by developing mutually exclusive and specific definitions of the four hypothesized humor dimensions […] We then generated a pool of items, sampling as many aspects of each hypothesized dimension as possible." (p. 55). After several rounds of generating, testing, and refining items (stage 4), 32 items were selected for the HSQ (stage 5).

1.2. *Construct validity of the HSQ*

Validity is the most important psychometric property, yet it often remains underexamined (cf. Ruch and Heintz 2014). Without information on (construct) validity, however, the degree to which an instrument is actually measuring what it is intended to measure is unknown. Relating one's instrument to other variables (such as well-being and personality) yields insights into the nomological network, yet understanding the essence of the constructs is indispensable (Cronbach and Meehl 1955). Validity can and should be assured in the process of test construction, for example, by taking a construct-based scale construction approach (Jackson 1970). However, whether this approach was successful or not still needs to be empirically tested, as potential shortcomings can occur at several stages, for example, when defining the construct to be measured, when operationalizing the construct (i.e., the translation from theory to measurement), or when analyzing and refining the psychometric properties of the measure.

One way to test construct validity is the multitrait-multimethod (MTMM) analysis. The MTMM analysis involves measuring multiple traits (MT) with multiple methods (MM), and the pattern of their intercorrelations yields conclusions about the construct validity (i.e., convergent and discriminant validity) of an instrument (Campbell and Fiske 1959). Methods can involve, for example, self- and peer-reports, aggregated states, behaviors, subtests, and contents. The aim of the MTMM analysis is to separate the trait-specific variance from the method-specific variance (and error variance), that is, to see to what extent the various methods are measuring the same constructs. In the present study, we use the definitions (stage 2), the aggregated construct descriptions (henceforth referred to as "constructs", stage 3), and the HSQ (stage 5) as provided by Martin et al. (2003) as three different indicators of the four humor styles (i.e., three methods and four traits in the MTMM framework). This MTMM analysis reveals how the conceptualizations of the humor styles converge with their actual measurement and thus provides important insights into the construct validity of the HSQ. If it happens that the convergence is low, then the basic assumption that all indicators tap into the same constructs is wrong.

Convergent validity would be high if the definition, the construct, and the HSQ scale of a humor style correlated highly with one another, indicating that they indeed tap into the same construct. The higher the convergent validity, the stronger the support that the HSQ scales map to their hypothesized humor dimensions (as outlined in stages 2 and 3) and that the construct-based scale construction was successful in this respect. The higher the discriminant validity, the better distinguishable the four humor styles are within and across every indicator. Investigating discriminant validity is important because the definitions were developed as being "mutually exclusive and specific" (Martin et al. 2003, p. 55) and because there are usually medium-sized positive correlations between the HSQ affiliative and self-enhancing scales (e.g., Cann and Matson 2014; Galloway 2010; Martin et al. 2003; Ruch & Heintz, 2013).

Only a few studies so far investigated the construct validity of the HSQ beyond its well-studied factorial validity (i.e., a replicable four-factor structure) and nomological network (i.e., correlations with other variables such as personality and well-being; but see Ruch and Heintz 2013). Martin et al. (2003) reported in their construction article the relationships of the HSQ scales with four peer-reported HSQ items, which showed small to medium convergent validities and satisfying discriminant validities (except between the affiliative and self-enhancing humor styles). When the humor behaviors of the HSQ were more closely studied (daily frequencies or humor removed from its context), discriminant validity among the four scales decreased and small to large positive intercorrelations were found among all humor styles (Caird and Martin 2014; Ruch and Heintz 2015). In a revised version of the HSQ (using situations and responses as items), a three-factor structure with positive, aggressive, and self-defeating humor was proposed, that is, affiliative and self-enhancing were not distinguishable (Reff 2006).

1.3. *The present study*

The present study aims at further analyzing the construct validity of the HSQ, specifically the convergence with its conceptualization. To this end, we rephrased the definitions and construct descriptions of the four humor styles given by Martin et al. (2003; pp. 52–54) into self-report

statements. We first analyzed these statements together with the HSQ in an MTMM analysis using two different statistical approaches (correlations and structural equation modeling). Comparing each humor style across its three indicators (definition, construct, and HSQ scale) provides a test of the convergent validity of the HSQ, that is, to what extent the HSQ scales correspond to the conceptualizations of the humor styles. Comparing the four humor styles within and across the three indicators provides a test of discriminant validity, that is, to what extent the four humor styles can be distinguished from one another.

Drawing on the previous findings with the HSQ, we expected that the discriminant validity would be high between all humor styles except for the affiliative and self-enhancing humor styles. As the construct definitions of the humor styles were developed as being "mutually exclusive and specific" (Martin et al. 2003, p. 55), discriminant validity among the four definitions should be high. With regards to convergent validity, we would expect a convergence between all three indicators (i.e., the definitions, constructs, and HSQ scales) of the four humor styles.

Second, we tested how representative the HSQ is with regards to the humor style conceptualizations. Ideally, the HSQ would represent the definitions (stage 2) and construct descriptions (stage 3) of the humor styles to a large extent. This would underscore that the individual differences in the conceptualizations are still (mostly) reflected in the measurement of the humor styles; that is, the HSQ scales should be able to explain most of the reliable variance in the definitions and constructs. This was investigated by predicting the definitions and constructs with the HSQ scales in multiple regression analyses.

We expected that the HSQ affiliative scale would represent its definition, but to a lesser extent its construct descriptions. The latter contain elements of friendly teasing, which is missing in the HSQ scale, and making fun of oneself (in a self-accepting way) to put others at ease, which is represented with one item only in the HSQ scale. The HSQ self-enhancing and aggressive scales should reflect their constructs, but to a lesser degree their definitions. In contrast to the self-enhancing definition ("humor may be used to enhance the *self* in a way that is tolerant and non-

detrimental to others"; Martin et al. 2003, p. 52), the HSQ self-enhancing scale only contains

settings of being alone or in a negative mood (like being sad, depressed, and upset), and the benign

aspect regarding one's relationships is not apparent. A similar observation can be made for the HSQ

aggressive scale: Its definition entails the function of using humor to enhance oneself, while the

scale does not. For the HSQ self-defeating scale, we would expect it to mostly reflect its definition

and constructs, despite the fact that the aspect of hiding one's underlying negative emotions is

represented in only one HSQ item.

2. Methods

2.1. *Participants*

Of the 584 German-speaking participants that started the survey, 362 (62.0%) filled in all

the items and statements. Importantly, there were no significant differences in age, gender, and

education between those who did and those who did not finish the survey (all ps > .40). A total of

340 participants (20.3% men) with a median age of 24 ($M = 28.30$, $SD = 10.06$) ranging from 18 to

64 years provided valid responses in this study (eight participants were excluded because they

answered more than 12 items/statements per minute, nine showed inconsistent response patterns or

invariant responses, and five indicated an age below 18 years). Participants were primarily Swiss

(60.6%) and German (31.5%). Most participants were well-educated, with 41.8% being college or

university students, 31.2% having passed tertiary education, 20.0% having A-levels, and 5.9%

having an apprenticeship.

2.2. *Instruments*

2.2.1. *Humor Styles Questionnaire (HSQ; Martin et al. 2003).* The HSQ consists of 32 items

measuring four humor styles. Sample items are "I enjoy making people laugh" (affiliative), "Even

when I'm by myself, I'm often amused by the absurdities of life" (self-enhancing), "If I don't like

someone, I often use humor or teasing to put them down" (aggressive), and "I let people laugh at

me or make fun at my expense more than I should" (self-defeating). The instrument employs a

seven-point Likert scale from "totally disagree" (1) to "totally agree" (7). Internal consistencies (Cronbach's alpha) ranged from .77 to .81. The German version of the HSQ was developed using a translation back-translation procedure, and the psychometric properties and factor structure in the present study were comparable to the original English version (high factor similarity as evidenced by Tucker's phi values of .90–.97 for the four factors).

2.2.2. *Definitions of the four humor styles.* The definitions along the 2×2-conceptualization given by Martin et al. (2003, p. 52) were rephrased with as little modification as possible into four self-report statements: "I use humor to enhance my relationships with others in a way that is relatively benign and self-accepting" (affiliative), "I use humor to enhance myself in a way that is tolerant and non-detrimental to others" (self-enhancing), "I use humor to enhance myself at the expense and detriment of my relationships with others" (aggressive), and "I use humor to enhance my relationships with others at the expense and detriment of myself" (self-defeating). These statements were answered with a seven-point Likert scale from "totally disagree" (1) to "totally agree" (7).

2.2.3 *Construct descriptions of the four humor styles.* The construct descriptions of each humor style were extracted (Martin et al. 2003, pp. 52–54) and rephrased with as little modification as possible into 23 self-report statements (with five to seven descriptions per humor style). Sample statements are "I gently tease or playfully poke fun at others within my own group" (affiliative), "I tend to be frequently amused by the incongruities of life" (self-enhancing), "My humor is intended to belittle others, albeit often under the guise of playful fun" (aggressive), and "I allow myself to be the 'butt' of others' humor" (self-defeating). These statements were answered with a seven-point Likert scale from "totally disagree" (1) to "totally agree" (7). The 23 statements are listed in Table 4.

2.3 *Procedure*

The data were collected in an online survey (www.unipark.info). The HSQ was presented first, followed by the self-report statements of the definitions and then the construct descriptions of

the humor styles (with an alternating sequence of the four humor styles). Further variables were collected that are not relevant to the present study. Participants were recruited via several means, including mailing lists, social media platforms, and bulletins of different universities. They were offered a personalized feedback and/or course credit in psychology for their participation.

2.4 *Data analysis*

Two MTMM analyses were conducted to test the construct validity of the HSQ, both employing four traits (i.e., the four humor styles) and three methods (the definitions, constructs, and HSQ scales as indicators of the humor styles). The first MTMM analysis was based on Campbell and Fiske's (1959) classic approach of analyzing the intercorrelation matrix of the four traits (humor styles) × three methods (indicators). The second MTMM analysis employed the structural equation modeling framework, which allows to model traits and methods as latent factors and to separate the variance that is due to trait, method, and error (for an overview, see Eid et al. 2008). Specifically, a single-indicator correlated-traits correlated-methods minus one (CTC[M-1]) model with the HSQ as the reference method was computed (Eid 2000), using *R* (R Development Core Team 2012) and the *lavaan* package (Rosseel 2012). In addition, the correlations between the single construct descriptions and the HSQ scales were investigated. To test the representativeness of the HSQ, standard multiple regression analyses with the definitions and constructs of the humor styles as criteria and the HSQ scales as predictors were computed.

3. Results

3.1. *Classic multitrait-multimethod (MTMM) analysis*

In this classic MTMM analysis, convergent and discriminant validity is indicated by the patterns of intercorrelations of the four traits × three methods. Table 1 shows the means, standard deviations, and the MTMM intercorrelation matrix of the HSQ scales, the definitions, and the constructs of the four humor styles.

Insert Table 1 about here

As can be seen in Table 1, all values on the validity diagonals (marked in dark gray) were significant and ranged from $r = .14$ to $r = .79$ (small to large effects). The validity diagonals indicated a high convergent validity of the four HSQ scales with the constructs (shared variance between 32.5% for affiliative and 62.4% for self-defeating). The convergence of the HSQ scales with the definitions of the humor styles was large for self-defeating (49.0% shared variance), medium for affiliative and aggressive (17.6% and 12.3%, respectively), and low for self-enhancing (2.0%). The definitions and constructs of the humor styles converged with medium to large effects (22.1–39.7% of shared variance). Thus, the three indicators of the self-defeating humor style converged to a large extent, while the definitions and the HSQ scales of the other three humor styles overlapped to a small to medium extent.

Comparing the validity diagonals with the corresponding heterotrait-monomethod (HTMM) triangles allows studying the discriminant validity of the four humor styles. Eight correlations were at least as high in the HTMM triangles (marked in light gray in Table 1) as in the corresponding validity diagonal, indicating a lack of discriminant validity among the constructs involved. Four of these pertained to the affiliative, five to the self-enhancing, and two to the aggressive humor style. Across all indicators, the HTMM correlations between the affiliative and self-enhancing humor styles (rs > .41, ps < .001) were larger than the convergent validities of the self-enhancing definition with its construct and HSQ scale. Furthermore, these correlations approximated the validities of the affiliative construct with its definition and HSQ scale. Thus, the conceptualizations of the affiliative and self-enhancing humor styles could not be well distinguished from one another. The self-enhancing definition also showed a low discriminant validity to the aggressive and self-defeating definitions, as did the affiliative and self-defeating constructs. The HTMM correlations of the aggressive and self-defeating definitions and constructs (rs > .41, ps < .001) were higher than the corresponding validities of the aggressive humor style, indicating a low discriminant validity. The self-defeating humor style exhibited consistently higher convergent validities than HTMM correlations, supporting its discrimination from the other three humor styles.

Comparing the validity diagonals with the homologous heterotrait-heteromethod (HTHM) triangles also leads to conclusions on the discriminant validity of the four humor styles. The self-defeating and aggressive humor styles showed a good discriminant validity, as no HTHM correlations were higher than their validities. All of the 7 (of a total of 36) correlations that were higher in the HTHM triangles than in the validity diagonal involved the self-enhancing humor style, mainly its definition (five correlations). In addition, the affiliative humor style overlapped to a medium to large extent with self-enhancing ($.29 < r$s $< .50$, ps $< .001$) across all three methods.

The MTMM intercorrelation matrix also reveals information on the factor-pattern similarity (see Campbell and Fiske 1959), indicating the extent to which the correlations between the humor styles are similar across the three indicators. Comparing the three HTMM-triangles, the intercorrelations among the HSQ scales were somewhat similar to those among the definitions (.51) and the constructs (.56), while the latter two were most similar (.77). The lower similarity between the HSQ and the definitions and constructs might be due to the correlations between self-defeating and affiliative, self-enhancing and aggressive, and self-enhancing and self-defeating, respectively, which were close to zero for the HSQ scales, but significant and positive for the definitions and constructs. The HTHM-triangles (comparing the triangle above with the triangle below the validity diagonal) revealed a high similarity between the intercorrelations of the HSQ scales and definitions (.76), a medium similarity between the HSQ scales and the constructs (.59), and a low similarity between the definitions and constructs (.28). This low similarity was mainly due to a medium-sized positive relationship between the self-defeating construct and the self-enhancing definition, while the self-enhancing construct and the self-defeating definition were uncorrelated. Thus, the relationships between the four humor styles were similar for some of the three indicators, but not across all of them.

3.2. *Correlated-traits correlated-methods minus one (CTC[M-1]) mode*l

A CTC(M-1) model with four traits (humor styles) and three methods (definitions, constructs, and HSQ scales) with the HSQ as the reference method was computed (using the

maximum likelihood estimator). The model showed a modest fit, $\chi^2(39) = 253.72$, $p < .001$, CFI = .890, TLI = .814, RMSEA = .127, 90% CI [.113, .142], and SRMR = .096. Table 2 shows the standardized trait and method loadings, the reliability, consistency (trait variance), and method specificity (method variance) of the CTC(M-1) model.

Insert Table 2 about here

As can be seen in Table 2, convergent validity was highest for the self-defeating and aggressive humor styles, with traits explaining 55–85% and 22–76% of the variance, and methods only 0–16% and 4–13%, respectively. The affiliative and self-enhancing humor styles, by contrast, evidenced high trait loadings of their HSQ scale and constructs, but not of their definitions (consistency 23% and 4% and method specificity 16% and 66%, respectively). These findings regarding convergent validity were similar to the classic correlational analysis, with the aggressive humor style performing slightly better and the affiliative humor style slightly worse in the CTC(M-1) model.

The intercorrelations among the latent trait factors in the CTC(M-1) model are indicative of the discriminant validity of the humor styles across the three indicators. Table 3 shows the variance with standard errors and correlations of the latent trait factors (four humor styles) in the CTC(M-1) model.

Insert Table 3 about here

As can be seen in Table 3, the affiliative humor style correlated significantly and positively with all other humor styles, most highly with self-enhancing (46.2% shared variance). The aggressive humor style correlated positively with the self-enhancing and self-defeating humor style (small to medium effects). The correlation between the methods (definitions and constructs) was .47

($p = .020$). Overall, discriminant validity was given for all but the affiliative and self-enhancing humor styles.

To elucidate the relationship between the HSQ scales and constructs in more detail, Table 4 shows the correlations between the 23 construct descriptions and the four HSQ scales.

Insert Table 4 about here

As can be seen in Table 4, two of the five affiliative descriptions correlated highest with the HSQ affiliative scale (AF3 and AF5), while the others correlated most strongly with one of the other HSQ scales. Thus, the affiliative construct descriptions also involved self-enhancing, aggressive, and self-defeating elements, which can explain the significant and positive intercorrelations of the affiliative construct with the other humor styles (see Table 1). With regards to self-enhancing, four of the five construct descriptions correlated highest with the HSQ self-enhancing scale, yet they also showed medium to large intercorrelations with the HSQ affiliative scale. One description (SE1), by contrast, correlated negatively with the self-enhancing and positively with the HSQ aggressive scale. The aggressive and self-defeating construct descriptions all correlated highest with their intended HSQ scale (medium to large effects), and they showed consistent intercorrelations with the HSQ self-defeating and aggressive scales (small to medium effects), respectively. In addition, three of the aggressive and self-defeating construct descriptions correlated positively with the HSQ affiliative scale (small effects). Thus, the fit of some of the affiliative and self-enhancing construct descriptions to the corresponding HSQ scales was low, while it was supported for all aggressive and self-defeating construct descriptions.

3.3 *Conceptual representativeness of the HSQ*

The conceptual representativeness of the HSQ scales would be supported if they reflected most elements of their definitions and constructs (approaching their internal consistencies in Table 1 and the estimated reliabilities in Table 2). The results of standard multiple regression analyses

with the definitions and constructs of the four humor styles as criteria and with the HSQ scales as predictors are presented in Table 5. To evaluate the unique contribution of each humor style in explaining the definitions and constructs, squared semi-partial correlations (sr^2) were computed.

Insert Table 5 about here

Table 5 shows that the HSQ explained overall 13.2% (self-enhancing definition) to 49.6% (self-defeating definition) of the variance in the definitions of the four humor styles (large effects). The homologous HSQ scales explained most variance in predicting the definitions (medium effects for affiliative and aggressive and a large effect for self-defeating) with the exception of the self-enhancing definition. The latter was mainly explained by the HSQ affiliative and self-defeating scales (small to medium effects). Small effects were also observed for the HSQ self-enhancing scale predicting the affiliative definition and for the HSQ affiliative and self-defeating scales predicting the aggressive definition. Compared with the reliability coefficients from the CTC(M-1) model, the four HSQ scales explained from 18.9% (self-enhancing) to 90.2% (self-defeating) of the estimated reliable variance in the definitions (51.3% of the affiliative and 68.5% of the aggressive definition).

Regarding the constructs, the HSQ explained 47.1% (affiliative) to 66.2% (self-definition) of the variance (large effects). In each case, the homologous HSQ scale explained the most variance. In addition, the affiliative construct was explained by the self-enhancing (small effect) and self-defeating (medium effect) HSQ scales. The self-enhancing construct was also explained by the HSQ affiliative and aggressive scales (small effects). Lastly, the HSQ self-defeating and aggressive scales explained a small amount of additional variance in the aggressive and self-defeating constructs, respectively. Compared with the internal consistencies (which establish a lower bound of reliability), the four HSQ scales explained from 64.4% (aggressive) to 100% (self-enhancing) of the reliable variance in the constructs; the same results were obtained when using McDonald's omega as an alternative measure of reliability, with the exception of self-enhancing that dropped to 90.9% of explained variance. Compared with the reliability coefficients from the CTC(M-1) model,

the four HSQ scales explained from 57.9% (aggressive) to 77.9% (self-defeating) of the reliable variance in the constructs. Thus, the HSQ scales represented several elements of their definitions and constructs and explained on average 57.2% of the reliable variance in the definitions and on average 65.6% (estimated reliabilities) to 79.4% (internal consistencies) in the constructs.

4. Discussion

This study aimed at relating the conceptualizations (stages 2 and 3 in the construction of the HSQ) to the measurement of the four humor styles (1) by investigating the convergence of the HSQ with its conceptualizations in MTMM analyses, and (2) by testing to what extent the HSQ represents its definitions and construct descriptions (as provided by Martin et al. 2003). With regards to the MTMM analyses, the convergence of the HSQ self-defeating scale with its conceptualization (definition and construct) was supported. For the other three HSQ scales, convergent validity was supported in relation to their constructs. By contrast, the definitions only partly converged with the HSQ affiliative and aggressive scale, and there was very little convergence with the HSQ self-enhancing scale. Discriminant validity was mostly supported for the aggressive and self-defeating humor styles, while the affiliative and self-enhancing humor styles were hard to distinguish from one another, especially with regard to their definitions and constructs. It was remarkable that the self-enhancing definition correlated lowest with its corresponding HSQ scale (and higher with the three others), so there was a clear mismatch between the definition and the measurement of this humor style.

Analyzing the single construct descriptions revealed that the affiliative construct descriptions spread across all HSQ scales, and the self-enhancing descriptions mainly correlated with the affiliative and self-enhancing HSQ scales (except for one statement). The aggressive and self-defeating construct descriptions consistently correlated with both the HSQ aggressive and self-defeating scales and correlated most strongly with their corresponding HSQ scale.

With regards to the conceptual representativeness of the HSQ, about two thirds of the reliable variance in the definitions and constructs of the four humor styles could be explained by the

HSQ. This empirically supports our content-based observation that some aspects of the humor style constructs were not reflected in the HSQ, and thus people's responses to the conceptualization (definitions and constructs) of the humor styles differed somewhat from their responses to the HSQ scales.

4.1. *Explanatory approaches*

Relating the conceptual framework of the humor styles to the contents of the HSQ scales, it seems that the initial 2×2 conceptualization (enhancing oneself vs. others $\times$ benevolent vs. detrimental) has been transformed into one dominant positive (affiliative and self-enhancing) vs. negative (aggressive and self-defeating) dimension. The enhancing oneself vs. others dimension has rather become a distinction between directing humor at others (affiliative and aggressive) and a more self-related kind of humor, involving humor appreciation and production independent of others (self-enhancing) or directing humor at oneself while being with others (self-defeating). Such changes might occur during the process of test construction, yet the original authors have not explicitly acknowledged them (and even listed the humor style definitions in the abstract of their construction article). Thus, the humor styles seem to have shifted away from their initial functions and now assess mainly adaptive vs. maladaptive uses of humor. This could – for example – explain why the aggressive humor style, which was initially conceptualized to aim at enhancing oneself, has been found to be either uncorrelated (e.g., Martin et al. 2003; Ruch and Heintz 2013) or negatively correlated with self-esteem (e.g., Edwards and Martin 2010; Galloway 2010).

Furthermore, Martin et al. (2003) acknowledged that "the distinction between potentially benign and deleterious uses of humor is one of degree, rather than a dichotomy" (p. 52), yet the HSQ items mainly seem to contain purely positive or negative aspects. This might also explain why the mixed construct descriptions of the "positive" humor styles (like AF1 and SE1, see Table 4) correlated highest with the HSQ aggressive humor style. In addition, this could explain why the HSQ scales showed a better discriminant validity than their definitions and constructs. This seems especially noteworthy as Martin et al. (2003) stated that "[w]e began by developing mutually

exclusive and specific definitions of the four hypothesized humor dimensions" (p. 55), yet in the present study, some of these definitions had a low discriminant validity (especially affiliative and self-enhancing).

A possible reason for the present findings might lie in the scale construction of the HSQ. Taking Jackson's (1970) system as a guideline, Martin et al. (2003) succeeded in obtaining reliable and rather distinguishable HSQ scales. The convergent validity to the 2×2 conceptualization and the conceptual representativeness, however, seem to have suffered a bit along the way (although they tried to avoid this). As for example Clark and Watson (1995) argue, focusing on obtaining reliable scales (especially by maximizing Cronbach's alpha as a measure of internal consistency) might lead to a lack of (construct) validity, as important facets of the construct might not be incorporated within the final scale (the classic "attenuation paradox"; Loevinger 1954). Another explanation could be that non-humorous elements (such as evaluations, situations, and moods) were added to the HSQ items that made the concepts more different than they actually are. This would be in accord with the finding that the differentiation between the HSQ scales was more driven by these non-humorous elements than by the humorous contents (Ruch and Heintz 2015).

It is important to note here that a perfect match between the construct conceptualization and its measurement is desirable, yet often not realistic due to the empirical scale refinement (for a thorough treatment of the process of item creation, see Angleitner et al. 1986). However, mismatches and larger deviations from the initial conceptualization should not occur, as then the measured constructs can no longer be interpreted along the theory or model they were derived from, but only in terms of their item contents (i.e., they no longer provide any surplus meaning; cf. MacCorquodale and Meehl 1948). Then either an adaptation or a new theory / model for the measurement is needed, or the constructs need to be investigated in more detail (e.g., in terms of content validity).

4.2. *Limitations*

One limitation of the present study is that it only involved self-reports. Employing other modes of data collection, such as peer-ratings and observations, would allow to test additional aspects of the construct validity of the HSQ, and would – from a psychometric point of view – help to facilitate validity judgments (as the correlation among the methods would be lower; see Carretero-Dios et al. 2011 for an example of such an MTMM analysis of a humor instrument). Second, the humor style definitions were rather global and only consisted of one statement each, which is suboptimal from a psychometric point of view. Specifically, convergent validity with the definitions might be underestimated and the discriminant validity overestimated, because there is likely more error variance than in aggregated statements. However, as our aim was to compare the HSQ with its conceptualizations, we did not want to change or add anything to the definitions as provided by Martin et al. (2003). A similar reasoning applies to the affiliative and self-enhancing constructs, which had low internal consistencies. Importantly, however, these issues (a) do not change the overall pattern of the correlations, and (b) are unlikely to have had a strong impact on our findings, as the theoretically possible correlations were still large ($\geq .44$). At the same time, we might overestimate convergent and underestimate discriminant validity, as we strictly used the statements made by the authors of the HSQ (Martin et al. 2003), which one would expect to be rather consistent. Third, the self-report statements of the conceptualizations were rather capturing different kinds of humor use, while the HSQ contains many items referring to the frequency of the humor use. These differences might have lowered the convergent and discriminant correlations obtained in the present study. Fourth, mainly well-educated females from Switzerland and Germany took part in the study. Thus, the conclusions might not be generalizable across the general population and other cultures and languages, and replications of the findings with divergent samples and particularly with the original HSQ version are desirable.

4.3. *Future directions*

The present study points out important research areas for future studies of the HSQ (and humor measures in general). First, it underscores the relevance of investigating the validity of

instruments (best during their construction). Only if an instrument turns out to be valid, researchers and practitioners will know what they actually measure and how their results can advance science and our understanding of the theories and constructs involved. Second, it points to the need of multi-methodological studies and alternative operationalizations of constructs to increase our knowledge beyond single questionnaires, like behavior observations, peer-reports, expert ratings, behavioral acts, diary studies, experimental settings, and so on. For example, deriving construct descriptions and definitions directly from the relevant theories (stage 1 in the construction of the HSQ) or employing experts ratings instead of self-reports by participants would be another approach to test the content and construct validity of the HSQ (for examples of content validation of humor instruments, see Carretero-Dios et al. 2009; Delgado-Rico et al. 2012). The statements used in the present study were based on Martin et al.'s (2003) construction article, but additional sources are conceivable (e.g., experts generating statements along the definitions of the humor styles, a prototype / act-frequency approach to derive behaviors indicative of the humor styles).

In general, we think that it is necessary to conceptually and theoretically link the humor styles to the construct or purpose under study to be able to draw valid and interpretable conclusions and to meaningfully advance the nomological network of the humor styles. Assessing humor behaviors or their functions directly would allow for an inherently valid assessment of humor styles. If using the HSQ, the warranted interpretations and conclusions of the findings should be carefully discussed (see also Ruch and Heintz 2013).

4.4. *Conclusions*

In summary, the present study showed that there was no strict convergence between the three indicators of all four humor styles: The indicators of the self-defeating humor styles converged well, while a lower convergence or even mismatches (self-enhancing definition, affiliative construct descriptions) were observed for the other three humor styles. The discrimination between the humor styles was mostly supported, with the exception of the affiliative and self-enhancing definitions and constructs. Furthermore, the HSQ scales explained about two-

thirds of the reliable variance in the definitions and constructs, so they represent several conceptual elements, while others are missing. Overall, then, the conceptual convergence and representativeness of the HSQ in terms of its conceptual foundations has to be questioned. What are implications of these findings? The constructs of the humor styles underlying the HSQ require further research before the HSQ can be validly used to advance our knowledge of humor styles and everyday functions of humor. If the present results can be replicated, then either the conceptualization underlying the HSQ needs to be adjusted or newly developed, or the HSQ does. Only then will we be able to understand the meaning and implications of the nomological network of the HSQ, that is, its relationships to other constructs (like well-being, personality, and social behavior), and to implement suitable applications (such as humor or well-being trainings).

5. References

Angleitner, Alois, Oliver P. John & Franz-Josef Löhr. 1986. It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In Alois Angleitner & Jerry S. Wiggins (eds.), *Personality assessment via questionnaires*, 61–108. Berlin: Springer.

Caird, Sara & Rod A. Martin (2014). Relationship-focused humor styles and relationship satisfaction in dating couples: A repeated-measures design. *Humor: International Journal of Humor Research* 27. 227–247.

Campbell, Donald T. & Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56. 81–105.

Cann, Arnie & Chelsea Matson. 2014. Sense of humor and social desirability: Understanding how humor styles are perceived. *Personality and Individual Differences* 66. 176–180.

Carretero-Dios, Hugo, Michael Eid & Willibald Ruch. 2011. Analyzing multitrait-mulitmethod data with multilevel confirmatory factor analysis: An application to the validation of the State-Trait Cheerfulness Inventory. *Journal of Research in Personality* 45. 153–164.

Carretero-Dios, Hugo, Cristino Pérez & Gualberto Buela-Casal. 2009. Content validity and metric properties of a pool of items developed to assess humor appreciation. *The Spanish Journal of Psychology* 12. 773–787.

Clark, Lee A. & David Watson. 1995. Constructing validity: Basic issues in objective scale development. *Psychological Assessment* 7. 309–319.

Cronbach, Lee J. & Paul E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52. 281–302.

Delgado-Rico, Elena, Hugo Carretero-Dios & Willibald Ruch. 2012. Content validity evidences in test development: An applied perspective. *International Journal of Clinical and Health Psychology* 12. 449–460.

Edwards, Kim R. & Rod A. Martin. 2010. Humor creation ability and mental health: Are funny people more psychologically healthy? *Europe's Journal of Psychology* 6(3). 196–212.

Eid, Michael. 2000. A multitrait-multimethod model with minimal assumptions. *Psychometrika* 65. 241–261.

Eid, Michael, Fridtjof W. Nussbeck, Christian Geiser, David A. Cole, Mario Gollwitzer & Tanja Lischetzke. 2008. Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods* 13. 230–253.

Freud, Sigmund. 1928. Humour. *The International Journal of Psychoanalysis* 9. 1–6.

Galloway, Graeme. 2010. Individual differences in personal humor styles: Identification of prominent patterns and their associates. *Personality and Individual Differences* 48. 563–567.

Jackson, Douglas N. 1970. A sequential system for personality scale development. In Charles D. Spielberger (ed.), *Current topics in clinical and community psychology*, vol. 2, 61–96. New York: Academic Press.

Loevinger, Jane. 1954. The attenuation paradox in test theory. *Psychological Bulletin, 51*. 493–504.

MacCorquodale, Kenneth & Paul E. Meehl. 1948. On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*, 95–107.

Martin, Rod A., Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray & Kelly Weir. 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* 3. 48–75.

R Development Core Team. 2012. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Reff, Robert C. 2006. *Developing the Humor Styles Questionnaire–Revised: A review of the current humor literature and a revised measure* (Doctoral dissertation). Pullman, WA: Washington State University dissertation.

Rosseel, Yves. 2012. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48(2). 1–36.

Ruch, Willibald & Sonja Heintz. 2013. Humour styles, personality, and psychological well-being: What's humour got to do with it? *European Journal of Humour Research* 1(4). 1–24.

Ruch, Willibald & Sonja Heintz. 2014. Test measurements of humor. In Salvatore Attardo (ed.), *Encyclopedia of humor studies*, vol. 2, 759–761. Thousand Oaks: Sage.

Ruch, Willibald & Sonja Heintz. 2015. *Extracting vs. replacing the core content in questionnaire items: An exemplification with the Humor Styles Questionnaire.* Switzerland: University of Zurich (Unpublished manuscript).

Zillmann, Dolf. 1983. Disparagement humor. In Paul E. McGhee & Jeffrey H. Goldstein (eds.), *Handbook of humor research: Vol. 1. Basic issues*, 85–107. New York: Springer.

Table 1.

Means, standard deviations, and the multitrait-multimethod intercorrelation matrix of the Humor Styles Questionnaire (1), the definitions of the four humor styles (2), and the aggregated construct descriptions (constructs) of the four humor styles (3).

		M	*SD*	(1) Humor Styles Questionnaire				(2) Definitions				(3) Constructs			
				AF	SE	AG	SD	AF	SE	AG	SD	AF	SE	AG	SD
(1)	AF	5.28	1.13	(.91)											
	SE	4.48	1.07	.41***	(.87)										
	AG	3.36	0.90	.28***	-.02	(.74)									
	SD	3.08	1.11	.03	.00	.20***	(.87)								
(2)	AF	5.14	1.14	.42***	.29***	.09	.08	–							
	SE	4.10	1.46	.29***	.14**	.19***	.20***	.47***	–						
	AG	2.14	1.09	.00	-.02	.35***	.28***	.07	.23***	–					
	SD	2.65	1.43	-.02	-.03	.14*	.70***	.14**	.17**	.41***	–				
(3)	AF	4.76	0.92	.57***	.38***	.28***	.35***	.47***	.42***	.15**	.27***	(.65)			
	SE	4.12	0.93	.50***	.70***	.25***	.06	.37***	.30***	.16**	-.04	.57***	(.58)		
	AG	2.28	0.94	.11	-.09	.69***	.29***	.08	.24***	.47***	.29***	.23***	.24***	(.80)	
	SD	3.21	1.01	.14*	.03	.34***	.79***	.17**	.31***	.36***	.63***	.45***	.20***	.50***	(.82)

Note. $N = 340$. AF = affiliative, SE = self-enhancing, AG = aggressive, SD = self-defeating. White = Heterotrait-heteromethod (HTHM) triangles, Light gray = Heterotrait-monomethod (HTMM) triangles, dark gray = validity diagonals. Cronbach's alpha in parentheses.

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

Table 2.

Standardized trait and method loadings, reliability, consistency, and method specificity of the correlated-traits correlated-methods minus one (CTC[M-1]) model with four trait (four humor styles) and three method factors (Humor Styles Questionnaire [HSQ], definitions, and aggregated construct descriptions [Construct]).

	Trait loading	Method loading	Reliability	Consistency	Method specificity
Affiliative					
HSQ	.68[+]		.46	.46	
Definition	.48***	.40[+]	.39	.23	.16
Construct	.81***	.29[+]	.74	.66	.08
Self-enhancing					
HSQ	.76[+]		.58	.58	
Definition	.20***	.81**	.70	.04	.66
Construct	.93***	.27***	.94	.86	.07
Aggressive					
HSQ	.75[+]		.56	.56	
Definition	.47***	.20**	.26	.22	.04
Construct	.87***	.36***	.89	.76	.13
Self-defeating					
HSQ	.92[+]		.85	.85	
Definition	.74***	.07	.55	.55	.00
Construct	.83***	.40***	.85	.69	.16

Note. $N = 340$. [+] = loading set to 1. Reliability = consistency plus method specificity. Consistency = squared trait loadings. Method specificity = squared method loadings.

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

Table 3.

Variance with standard errors (SE) and correlations of the latent trait factors (four humor styles) in the correlated-traits correlated-methods minus one (CTC[M-1]) model.

Traits	Variance	*SE*	(1)	(2)	(3)	(4)
(1) Affiliative	0.59	0.09	1.00			
(2) Self-enhancing	0.66	0.09	.68***	1.00		
(3) Aggressive	0.44	0.07	.14*	.14*	1.00	
(4) Self-defeating	1.05	0.10	.32***	.05	.38***	1.00

Note. $N = 340$.

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

Table 4.
Correlations of the construct descriptions with the Humor Styles Questionnaire (HSQ) scales.

		HSQ AF	HSQ SE	HSQ AG	HSQ SD
AF1	I gently tease or playfully poke fun at others within my own group.	.37***	.08	.40***	.13*
AF2	I am able to gently poke fun at my own faults and I do not take myself too seriously.	.40***	.43***	.10	.12*
AF3	I tend to say funny things, to tell jokes, and to engage in spontaneous witty banter to amuse others, to facilitate relationships, and to reduce interpersonal tensions.	.63***	.34***	.24***	.22***
AF4	To put others at ease, I am likely to engage in self-deprecating humor, saying funny things about myself and not taking myself overly seriously, while maintaining a sense of self-acceptance.	.13*	.14*	.17**	.56***
AF5	I have an essentially non-hostile, tolerant use of humor that is affirming of myself and others.	.38***	.29***	-.06	-.02
SE1	I derive some inner amusement and pleasure from observing or imagining the ignominious defeat of my adversaries.	-.03	-.15**	.36***	.06
SE2	I have a generally humorous outlook on life.	.61***	.70***	.11*	-.02
SE3	I tend to be frequently amused by the incongruities of life.	.33***	.44***	.16**	.00
SE4	I maintain a humorous perspective even in the face of stress or adversity.	.44***	.70***	.07	-.01
SE5	I regulate negative emotions through humorous perspective-taking.	.31***	.68***	.03	.15**
AG1	Denigrating, disparaging, excessively teasing, or ridiculing others enhances me.	-.06	-.14**	.38***	.14*
AG2	My humor is intended to belittle others, albeit often under the guise of playful fun.	.01	-.08	.54***	.23***
AG3	I use sarcasm, teasing, ridicule, derision, "put-down", or disparagement humor.	.15**	-.10	.58***	.19***
AG4	I use humor to manipulate others by means of an implied threat of ridicule.	-.04	-.08	.42***	.20***
AG5	I tend to express humor without regard for its potential impact on others (e.g., sexist or racist humor).	.17**	-.01	.55***	.22***
AG6	I compulsively express humor in which I find it difficult to resist the impulse to say funny things that are likely to hurt or alienate others.	.14*	.03	.47***	.27***
SD1	I use self-disparaging humor excessively.	.05	-.04	.32***	.65***
SD2	I attempt to ingratiate myself or gain the approval of others by doing or saying funny things at my own expense.	.09	.03	.20***	.69***
SD3	I engage in humorous behavior as a means of repressing my underlying feelings, in order to maintain the acceptance of others.	.01	-.04	.20***	.46***
SD4	I tend to engage in humorous behavior as a means of hiding my underlying negative feelings or avoiding dealing constructively with problems.	.03	-.03	.25***	.45***
SD5	I laugh along with others when being ridiculed or disparaged.	.12*	.03	.23***	.43***
SD6	I allow myself to be the ''butt'' of others' humor.	.24***	.12*	.25***	.43***
SD7	I attempt to amuse others by doing or saying funny things at my own expense as a means of ingratiating myself or gaining approval.	.13*	.05	.24**	.74***

Note. $N = 340$. AF = affiliative, SE = self-enhancing, AG = aggressive, SD = self-defeating.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5.

Standard multiple regression analyses with the definitions and the aggregated construct descriptions (constructs) of the four humor styles as criteria and with the Humor Styles Questionnaire (HSQ) scales as predictors (squared semi-partial correlations reported).

	Definitions				Constructs			
	AF	SE	AG	*SD*	AF	SE	AG	*SD*
HSQ AF	.103	.047	.010	.001	.160	.026	.004	.003
HSQ SE	.016	.001	.001	.000	.029	.324	.002	.000
HSQ AG	.001	.006	.100	.000	.007	.037	.402	.026
HSQ SD	.006	.031	.043	.473	.097	.000	.023	.542
R	.45	.36	.42	.70	.69	.77	.72	.81
Total R^2	.200	.132	.178	.496	.471	.591	.515	.662
F for R^2	20.99***	12.70***	18.18***	82.31***	74.68***	120.85***	88.78***	163.92***

Note. $N = 340$. AF = affiliative, SE = self-enhancing, AG = aggressive, SD = self-defeating, R = multiple correlation.

*** $p < .001$.